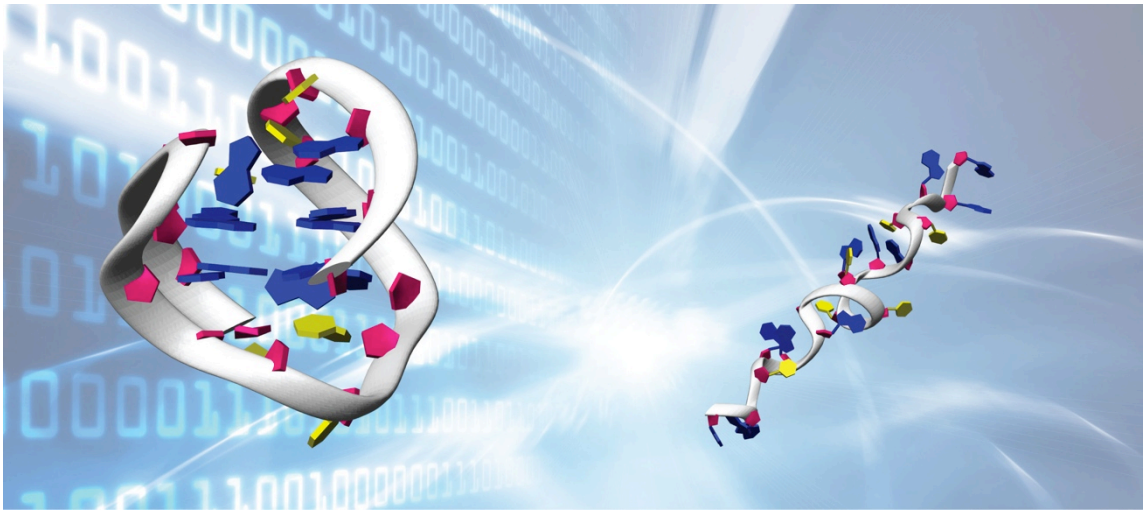
FRONT MATTER



ENCODING CANONICAL DNA QUADRUPLEX STRUCTURE

S. A. Dvorkin, A. I. Karsisiotis, M. Webba da Silva*

School of Pharmacy & Pharmaceutical Sciences, Biomedical Sciences Research Institute, Ulster University, Coleraine, BT52 1SA, UNITED KINGDOM.

* To whom correspondence should be addressed. Tel: 44 2870124009; Email: mm.webba-da-silva@ulster.ac.uk

ABSTRACT

The main challenge in DNA quadruplex design is to encode 3D structure into the primary sequence, despite its ubiquitous guanines. Here, we identify and detail structural elements describing all fourteen feasible canonical quadruplex scaffolds, and demonstrate their use in control of design. This work outlines a new roadmap for implementation of targeted design of quadruplexes for material, biotechnological and therapeutic applications.

INTRODUCTION

In the design of objects comprised of nucleic acids, it is crucial to avoid the presence of alternative molecular recognition motifs, since these may result in non-targeted architectures. At the same time, quadruplex DNA is characterized by the presence of repetitive segments of guanines, which have the potential for a great variety of alternative hydrogen bond alignments. The prediction of quadruplex architectural folds encoded in guanine-rich DNA sequence is an important challenge relevant for

understanding its significant regulatory roles(*1*) with implications on the evaluation of new therapeutic targets.(*2*) The current general consensus is that predicting, or controlling quadruplex folding is a mostly intractable problem. For example, biologically significant DNA sequences are often intrinsically polymorphic at *in vitro* conditions and respond to pH, cations or crowding conditions. Nevertheless, there is substantial interest in this theme due to the potential of these structures as functional materials,(*3, 4*) in templated-organization of materials,(*5*) as nanowires,(*6*) in catalysis,(*7*) and as therapeutics.(*8*) For example, a quadruplex-based nanomotor relying on a conformational switch between canonical quadruplex and duplex quadruplex resulted in approximately 5 nm displacement (*3*). Potential application of quadruplexes to nanoelectronics has been illustrated by measurements of currents greater than 100 pA in single molecule quadruplex wires over 100 nm long.(*6*) Direct imaging of quadruplex formation and their association with proteins has been enabled by constructing a quadruplex into a DNA origami scaffold.(*4*) These examples illustrate the significant utility of controlled design of quadruplexes. However, the design of quadruplexes is a complex problem also influenced by attributes of the self-assembly environment.

A major step in addressing this problem is to develop the ability to define its structural characteristics. Canonical quadruplexes are composed of a single strand containing at least four tracts of two or more guanines that form a stem. These are linked by three loops, d($G_nL_xG_nL_yG_nL_zG_n$); where n is the number of sequential guanines that form the stem, and L is the number (x, y, or z) of residues linking guanines in the stem. In the stem, a guanine from each of the four guanine-tracts engages in hydrogen bonds along its Hoogsteen and Watson-Crick edges to form a tetrad (also known as a quartet). In the stacked tetrads, each of the 2'-

deoxyguanosines adopts one of two conformational states that relate base to sugar through the glycosidic bond: the glycosidic torsion angle χ. These two non-overlapping ranges are described as either *syn* ($-90° \leq \chi \leq +90°$) or *anti* ($+90° \leq \chi \leq +180°$) conformation. This has implications for the depth of the grooves along the stem of the molecule. Instead of minor and major grooves found between each strand in double helical DNA, in quadruplexes the grooves can be described as *narrow (n)*, *medium (m)* and *wide (w)* as defined by the glycosidic bond conformation adopted by any of two base-paired 2'deoxyguanosines of stacked tetrads in the stem (Fig. 1a).

Throughout the publicly available literature, various names have been used to describe canonical quadruplex architecture. These include "chair-type", "basket-type", "(3+1) scaffold", "(2+2) scaffold", "form-1" and "form-2" telomeric scaffolds. Whilst these have had some merit for describing quadruplexes in the past, the structural diversity of quadruplexes can now be classified,(*9, 10*) and therefore systematic nomenclature can be developed. A canonical quadruplex can be denominated by a single descriptor containing the number of guanines in the stem, along with the type and relative direction of loops linking G-tracts of the stem: n(L1,L2,L3)- see Fig. 1. Direction of loop progression, L1 to L3, is described in relation to a frame of reference(*9*) beginning with the 5' end as the lower right corner of stem from the viewer's perspective. Clockwise progressing loops originating from this frame of reference are denoted with a '+' preceding the loop type, and loops progressing anticlockwise are denoted with a '-'preceding the loop type. Finally, the grooves of the quadruplex stem can be encoded in the description for lateral loops as "w" for *wide*, "n" for *narrow* or "m" for *medium grooves*, and denoted in subscript after the lateral loop designator. A lateral loop over a narrow groove is thus denoted l_n and, over a wide groove l_w.

This nomenclature does not directly address the position of the glycosyl bond conformation of the guanosines in the topology. Knowing the precise position of the glycosyl bond conformation of 2'-deoxyguanosines in a quadruplex stem is important for quadruplex design,(*11*) and exploitation of their physical properties.(*6, 12*) Eight possible groove type combinations exist which result from variations of glycosyl bond conformation of guanosines in the stem. Steps of glycosyl bond progressing through a strand of the stem can be either all the same glycosidic bond conformation, i.e. *syn-syn* or *anti-anti* throughout (Type 1), or two variations of mixed conformation.(*13, 14*) Individual steps between the two mixed conformations can either be alternating *syn-anti-syn*, (Type 3) or two identical conformations, either followed or preceded by the alternate conformation; i.e. *syn-syn-anti* or *anti-anti-syn* (Type 2). The positions of glycosyl bond conformations for Type 1, Type 2, and Type 3 quadruplex stem have been suggested, but some have yet to be demonstrated experimentally.(*11*)

Molecular recognition can be encoded in the DNA sequence through the formation of pseudo-planar hydrogen bonding alignments of guanosines. For this, knowledge of the glycosidic bond conformations to be adopted by guanosines throughout the stem of the targeted topology is required. This in turn enables segment length selection of residues in the primary sequence that will form loops in the final architecture. This approach has been previously applied in the design of a quadruplex architecture containing all three loop types.(*11*) However, to render the approach more generally applicable, the remaining questions are addressed herein. Specifically, the topologies for all feasible canonical quadruplexes are described, and this knowledge is applied to the design of canonical quadruplex architectures utilizing modified bases. We demonstrate that in the design of quadruplexes, the combination of lengths of loops is

dependent, not only on the arrangement of the groove widths, but also on the number of stacking tetrads of the quadruplex stem.

RESULTS

Propeller loops bridge parallel-stranded *syn*G-*syn*G-*anti*G grooves

To evaluate the feasibility of of propeller loop formation spanning grooves composed by parallel-stranded *syn*G-*syn*G-*anti*G, we designed the 3(d+pd) topology and determined its solution structure (Fig. 2), utilizing NMR spectroscopy. For experimental details, see Supplementary Material. The sequence d($G_3T_4G_3TG_3T_4G_3$) in 20 mM sodium adopts the target 3(d+pd) topology, see Fig. 2c-f. The right-handed Type 2 stem is composed of an equal number of *syn*G and *anti*G. The propeller loop spans a groove composed of parallel-stranded *syn*G-*syn*G-*anti*G, demonstrating that this structural motif is feasible in quadruplex structure.

The 5'-end of Type 3 stem adopts *syn* conformation

To evaluate whether a 3-stacked Type 3 quadruplex folds with a *syn*G in the 5'-end of the stem, we designed a DNA sequence targeting the 3($-l_wd+l_n$) topology (Fig. 3a), and determined its solution structure utilizing NMR spectroscopy. The DNA sequence d($G_3T_3G_3T_4G_3TG_3$) (5J05) folds into the 3($-l_wd+l_n$) topology in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 solution, Fig. 3c-e. Its right-handed Type 2 stem includes a *syn*G in its 5'-end. In order to evaluate whether formation of this motif is influenced by a base attached to the 5'-end of the stem, the DNA sequence d($TG_3T_3G_3T_4G_3TG_3$) (S093) was also found to adopt the same topology with a *syn*G in the 5'-end of the stem.

Encoding topology with conformationally locked 2'deoxyguanosine derivatives

To evaluate whether the formation of the 3(-l_wd+l_n) topology in d($G_3T_3G_3T_4G_3T_2G_3$) (S232) can be induced, dG7 in this sequence was replaced with an rG d(G_3T_3**rG**$G_2T_4G_3T_2G_3$) (S209). The sequence S209 folds into a single species in 100 mM sodium solution at pH 6.8 (Fig. 3e), and adopts the desired 3(-l_wd+l_n) topology.

Loop-length and number of stacking tetrads are interdependent in design

To evaluate the hypothesized interdependency of loop length combinations and number of stacking tetrads in the programming of quadruplex topologies, we designed 2-, 3-, and 4-stacked quadruplexes of the (-l_wd+l_n) topology with lateral loop variations as shown in Table 1, and structurally characterized them utilizing solution NMR spectroscopy (see Supplementary Material). Sodium solutions were selected since they favor formation of Type 3 stem.(*15*)

The solution structures of DNA sequences d($G_4T_3G_4T_4G_4A_2G_4$) (5J6U), d($G_4T_2G_4T_4G_4A_2G_4$) (2M6W), d($G_3T_2G_3T_4G_3TG_3$) (2M6V), d($G_2T_3G_2T_4G_2T_2G_2$) (5J4W), and d($G_2T_3G_2T_4G_2T_3G_2$) (5J4P) were determined and details are presented in Fig. 4. All sequences form right-handed quadruplexes with a *syn*G at the 5'-end of the stem, and alternation of glycosidic bonds along guanosine segments. The 4-thymine diagonal loop adopts a very similar structural environment in all of them, with lateral loop residues stacking onto the stem. The third thymine of the loop stacks onto the first (5'-end) *syn*G of the stem. The close proximity of these two residues was used for initial identification of successful formation of the topology for those sequences that folded successfully in solution and adopt the n(-l_wd+l_n) topology.

For 4-stacked architectures, the data suggests that a difference of more than one residue between first and third loop prevents folding (Table 1). Indeed, combinations of loops d($G_4T_2G_4T_4G_4TG_4$) (S036), d($G_4T_2G_4T_4G_4A_2G_4$) (2M6W),

d($G_4T_2G_4T_4G_4T_3G_4$) (S069), d($G_4A_2G_4T_4G_4T_3G_4$) (S067), d($G_4T_3G_4T_4G_4A_2G_4$) (5J6U), d($G_4T_3G_4T_4G_4T_3G_4$) (S080), d($G_4T_4G_4T_4G_4T_4G_4$) (201D), but not d($G_4T_3G_4T_4G_4AG_4$) (S064) and d($G_4T_4G_4T_4G_4A_2G_4$) (S066), are able to form 4(-l_wd+l_n).

When inducing 3-stacked quadruplexes, the probability of forming a propeller loop with combinations of shorter loops is higher.(*15, 16*) For 3-stacked architectures, combinations of short loops d($G_3T_2G_3T_4G_3AG_3$) (S025), d($G_3T_2G_3T_4G_3A_2G_3$) (S087) form multiple species in solution. Just as for the 4(-l_wd+l_n), only combinations in which the number of residues of both loops do not differ by more than a single residue were successful, i.e., d($G_3T_2G_3T_4G_3T_3G_3$) (S089), d($G_3A_2G_3T_4G_3T_3G_3$) (S088), d($G_3T_3G_3T_4G_3A_2G_3$) (S090), and d($G_3T_3G_3T_4G_3T_3G_3$) (S231) all adopt the 3(-l_wd+l_n) topology. However, in combinations involving 4-residue -l_w loops, d($G_3T_4G_3T_4G_3T_2G_3$) (S174), d($G_3T_4G_3T_4G_3T_3G_3$) (S175), and d($G_3T_4G_3T_4G_3T_4G_3$) (S038a), multiple species are observed.

Sequences that form 3(-l_wd+l_n) exist in which the number of residues of both loops differ by more than a single residue, including d($G_3T_3G_3T_4G_3AG_3$) 5J05 and its analogue d($TG_3T_3G_3T_4G_3AG_3$) S093. The analogue to 5J05 with a thymine instead of an adenine in the +l_n loop, d($G_3T_3G_3T_4G_3TG_3$) S029, forms multiple architectures in solution. Inspection of the 5J05 solution structure here determined suggests that the stabilization of this architecture may be due to stacking of the single adenine of the +l_n loop onto its subsequent *anti*G in the stem.

Short loops such as in d($G_2T_2G_2T_4G_2TG_2$) S166 and d($AG_2T_2G_2T_4G_2T_2G_2$) S230 failed to form 2(-l_wd+l_n), resulting in multiple species. However, with 2 residues per lateral loop, d($G_2T_2G_2T_4G_2T_2G_2$) S169 resulted in a single species but could not be characterized as 2(-l_wd+l_n). Just as for the 4(-l_wd+l_n) and 3(-l_wd+l_n), combinations in

which the number of residues of both loops do not differ by more than a single residue, d($G_2T_2G_2T_4G_2T_3G_2$) S172, d($G_2T_3G_2T_4G_2T_2G_2$) 5J4W, and d($G_2T_3G_2T_4G_2T_3G_2$) 5J4P, adopted the desired 2(-l_wd+l_n) topology. Like the 3(-l_wd+l_n), but in contrast to 4(-l_wd+l_n) sequences, d($G_2T_3G_2T_4G_2TG_2$) S167 and d($G_2T_4G_2T_4G_2TG_2$) S168 form 2(-l_wd+l_n) despite having a 2-residue difference between the opposing lateral loops.

An attempt to fold loop combination 2M6V into 3(-l_wd+l_n) proved unsuccessful. Instead, it adopted 2(-l_wd+l_n); with a 4-residue -l_w and a 2-residue +l_n; see Fig. 4. In place of the designed 2-residue -l_w, the 2(-l_wd+l_n) has a 4-residue loop that incorporates guanosines of the first and second G-rich segments. Thus, the energetically favored architecture still contains the expected diagonal loop formed by 4-residues, but also includes a longer (4-residue) loop for -l_w. Whilst the guanosine of the first G-rich segment forms a *syn*G:T mismatch, the guanosine of the second G-rich segment stacks perfectly onto its preceding *anti*G of the stem. This stabilization motif is also observed for 2-residue (2M6W), 3-residue (5J6U, 5J4W), and 4-residue (2KF8, 2KKA in Table 1) -l_w loops that proceed from a *syn*G to an *anti*G of the stem. The G6:T mismatch appears to be an essential stabilizing factor, since it is absent in the analogous DNA sequences that do not fold (S025).

The -l_w loop in 2- and 4- but not 3-stacked (-l_wd+l_n) can be stabilized by 4-residues. However, 3(-l_wd+l_n) can be stabilized by a 3-residue -l_w loop. This may be due to the fact that while in 2(-l_wd+l_n) and 4(-l_wd+l_n) the -l_w loop progresses from an *anti*G to a *syn*G of the stem, the reverse is true for 3(-l_wd+l_n). For a -l_w loop, a 3-residue *syn*G to *anti*G strand progression may be considered mechanically equivalent to a 4-residue *anti*G to *syn*G strand progression.

DISCUSSION

The challenge in design of quadruplexes is to determine the optimal sequence that reliably encodes for a given 3D structure. Therefore knowledge of structural motifs of the desired architecture is fundamental to this process. Twenty-six theoretical quadruplexes have been previously proposed,(*9, 11, 14*) however a fragment-based molecular mechanics approach applied to the derivation of 3-stacked quadruplexes predicted that only 14 of these are mechanically feasible.(*17*) Here, we utilize known and hypothesized glycosyl bond conformations to design 3D architectures in order to (i) provide experimental verification of all feasible glycosyl bond conformations in these 14 quadruplex topologies, (ii) demonstrate interdependency between loop-length combination and the number of stacking tetrads of quadruplexes.

In Fig. 5 the schematic representations of glycosidic bond conformations for fourteen feasible canonical quadruplex topologies are summarized. Only four of the nine Type 2 quadruplexes have been experimentally verified: 3(-p-l_w-l_n),(*18*) 3(-l_w-l_n-p),(*19*) 3(-pd+l_n),(*20*) and 3(+l_n+p+l_w).(*21*) In all of these, propeller loops bridge grooves of parallel-stranded *syn*G-*anti*G-*anti*G, demonstrating that the formation of the previously hypothesized(*11*), but as yet unobserved topologies 3(+l_nd-p), 3(-p-p-l_w), and 3(+l_n+p+p) are possible. The two remaining Type 2 topologies, 3(-pd+p) and 3(d+pd), have a propeller loop bridging a medium groove of parallel-stranded *syn*G-*syn*G-*anti*G. The propeller loop spanning this groove has been confirmed in this study, thus verifying the hypothesized glycosyl bond conformation. Therefore, this structural motif permits formation of the yet to be observed 3(-pd+p). Type 2 quadruplexes are here proposed to be restricted to 2- and 3-stacked topologies, since formation of propeller loops for 4-stacked canonical Type 2 quadruplexes is improbable.

All four Type 3 quadruplex topologies shown in Fig. 5 have been experimentally verified. For all 2-, and 4-stacked, the 5'-end of the stem is *syn*G: (+l_nd-l_w),(*22*) (-l_wd+l_n),(*23, 24*) (+l_n+l_w+l_n),(*25, 26*) and (-l_w-l_n-l_w).(*27*) In these architectures, more s*yn-anti* steps are present than the less stabilizing(*28*) *anti-syn* steps. However, a single 3-stacked structure exists with an *anti*G in the 5'-end of the stem: the human telomeric sequence d[$AG_3(T_2AG_3)_3$],(*29*) which adopts the topology 3(-l_wd+l_n) in 100 mM sodium solution at pH 6.8. Nonetheless, a *syn*G conformation has been suggested to be more stabilizing for a quadruplex stem.(*30*) Furthermore, in any 3-stacked Type 3 quadruplex, the number of *syn-anti* steps will be the same as *anti-syn*, regardless of the conformation adopted by the 5'-stem guanosine. All seven sequences designed to fold 3(-l_wd+l_n) topology were observed to adopt a *syn*G at the 5'-end of the quadruplex stem, thus demonstrating it to be the most stable conformation. Other Type 3 topologies may also adopt two alternative dispositions of glycosyl bond conformation. The preference for *anti*G at the 5'-end of the stem observed in the human telomeric sequence may outmatch the general preference for *syn*G through the additional stacking of triads or other mismatch alignments onto the Type 3 stem.

Individual loop length does not define successful design, it is their combination that does. For example, although four residues in the first, or one residue in the last loop each separately allow folding of (-l_wd+l_n), the combination of the two in a single sequence does not. The difference between lateral loops of 2- and 3(-l_wd+l_n) can be up to 2 residues, but only 1 residue for 4(-l_wd+l_n). Also, only 4(-l_wd+l_n) is able to fold with lateral loops of 1 or 2 residues.

Additionally, we determined that the number of stacking tetrads of the targeted topology predicates the optimal loop-length combination. Combinations of 2 and 3, or just 3-residue segments for the lateral loops enabled reliable assembly of any n(-

$l_w d+l_n$) topology. However, greater versatility is permitted for the design of 2($-l_w d+l_n$) architectures. These can fold with a $-l_w$ of 4 residues, in contrast to 3($-l_w d+l_n$), and can accommodate a larger difference of residues between lateral loop-lengths, in contrast to 4($-l_w d+l_n$).

Nucleobase substitutions can be utilized to induce quadruplex folding, however, not all site substitutions will result in well-folded species,(*31*) suggesting that strategies for their use require further development. Guanine to 8-bromoguanine substitution has been shown to induce the folding of *syn*G at the substituted site,(*21, 24, 32, 33*) and riboguanosine substitution leads to the folding of *anti*G.(*34, 35*) Our unsubstituted sequence d($G_3T_3G_3T_4G_3T_2G_3$) failed to adopt the expected 3($-l_w d+l_n$) topology (Table 1**)**, but substitution of a single *anti*G of 3($-l_w d+l_n$) by the conformationally locked rG enabled successful folding of the targeted topology. Although a number of antiG positions could have been selected for this substitution we chose the position at the end of the first loop. This exemplifies how identification of the glycosidic bond conformation of guanosines in the stem enables successful design.

The structural details of canonical quadruplex topologies provided give the explicit information required to design architectures for each topology. This will inform both experimental and theoretical approaches which will result in improved reliability and reproducibility in quadruplex design. Many principles of design are already established that make it possible to engineer the architectures described. For example, in canonical quadruplex architectures, positioning of loops (as first, second, and third) is fundamental in sequence design; i.e., reversing the loop-length sequence in the combination of loops does not result in the same topology.(*36*) Equally, choosing the nature of residues able to afford the desired loop combinations is significant. Here we utilized predominantly thymines, and less often adenines, while avoiding cytosines to

prevent formation of G:C base pairs. Inclusion of purines in loops may be used to facilitate tetrad stacking and formation of hydrogen bond alignments that should also stack onto tetrads of the quadruplex stem. Successful formation of Type 1 quadruplexes is favored in potassium, Type 2 and Type 3 in sodium solutions.(*15*) Potassium can also be used to fold Type 2 and Type 3 architectures.

Here we have utilized four residues to generate diagonal loops. MD simulation studies,(*28, 37*) and experimental studies,(*15*) show that formation of propeller loops is favored by 1 or 2 residues. Indeed, a combination of at least two such loop lengths is likely to be successful for Type 1 topology. Loops of 2 and 3 residues favor (l_n) over (l_w), while 4-residue loops favor (l_w).(*28*) However, lack of loop selectivity for single propeller loops and clockwise lateral loops present unique challenges in design. To mitigate these problems successful design can be further enabled by oligonucleoside modification strategies as demonstrated in this study. However, more general solutions can be sought from a greater understanding of the pre-folding states and folding pathways. This will allow for the development of strategies to control thermodynamic parameters modulating the kinetic routes for self-assembly.

In this study we describe the feasible canonical quadruplex topologies and structural requirements for their design. These scaffolds can be used to inform the design of quadruplex architectures, as well as fine tuning a desired topology for specific applications.

Potential applications in therapeutics and nanotechnology of DNA quadruplexes are varied and diverse. Quadruplex-forming DNA sequences are highly prevalent in mammalian and bacterial genomes where they have established regulatory roles. Small molecules that can stabilize these architectures may be utilized as therapeutics

and sensors.(*2*) The structural diversity library based on the set of 14 topologies described here can be interrogated by small-molecule combinatorial libraries for the discovery of leads against the quadruplex topologies represented. Conversely, the quadruplex structural diversity library can also be used to 'fish' for protein interaction partners in nuclear extracts in order to identify quadruplex binding proteins. Quadruplex topologies identified in this manner can then be further fine-tuned for the development of therapeutics, biomarkers, or other diagnostic purposes.

Knowledge of the 14 feasible canonical topologies allows for choice of quadruplex with structural characteristics appropriate for the loading of desired payloads. These may be existent or future drugs, for delivery to target locations through stimuli-responsive conformational change involving quadruplex structure and random single-stranded sequence. For example, superparamagnetic nanoparticles surface-functionalized with canonical quadruplexes carrying a payload for intracellular delivery can be thermally activated when subjected to an alternating magnetic field.(*38*) Design of quadruplexes can be tailored to create temperature-dependent unfolding (payload delivery) dependent on a range of desired temperatures. The approach can also be used for the construction of devices for targeted delivery to solid tumors.

Quadruplex nanodevices are known to be able to sense cations, small molecules, and proteins- reviewed in ref (*39*). Reversible quadruplex folding can therefore also be controlled through photoregulation or presence of specific cations. For example, in the presence of hemin, quadruplexes are able to catalyze hydrogen peroxide-mediated oxidation of 2,2′ -azinobis(3-ethylbenzothiazoline-6-sulfonic acid) to produce a color change which leads to chemiluminescence. Target recognition by the quadruplex is thus observed colorimetrically or through a fluorescence signal. The added ability to

tune quadruplex structure in this context will be a powerful tool for a variety of applications expected to include the identification of pathogens and development of diagnostics and biomarkers.

The availability of the feasible canonical topologies can be also used to inform the rational design of well-defined quadruplex nanowires(*40*). In principle, systematic design should allow for the modulation of electronic and photoelectronic properties of these materials due to their π-stacking system. This should ultimately enable the fine-tuning of properties that make these architectures good candidates for integrated multicomponent systems.

In summary, programming canonical quadruplex structure based on knowledge of the structural characteristics of the stem has been demonstrated. The complete description of all glycosyl bond angle conformations for canonical quadruplex topologies is presented for the first time. We also demonstrate the feasibility of alternative glycosyl bond conformations within a quadruplex topology. Optimal individual loop lengths have been shown to be dependent on the context of other loop types, as well as loop lengths, within the sequence. Finally, we have shown that the target number of stacking tetrads influences the combination of loop-lengths required for successful folding. The improved ability presented here to reliably engineer high-fidelity, topology-specific quadruplex architectures enables the exploitation of this molecule's unique potential.

MATERIALS AND METHODS

Sample preparation. Primary DNA sequences were synthesized, trityl off, by Eurogentec (Belgium), supplied desalted and lyophilized. They were subsequently purified by reverse phase HPLC, in ion-pair mode, on a Phenomenex Clarity Oligo-RP column (C18, 5 μm, 10 x 100 mm) using an acetonitrile gradient (5 - 40 % over 35 minutes) and a 100 mM TEAA buffer (pH 6.8). Typically, the recovered fractions were subsequently subjected to gel filtration using Sephadex® G-15 (Sigma Aldrich, Sweden), and finally three rounds of dialysis in water/sodium buffered solutions. After lyophilization, samples were re-suspended in concentrations of total sodium ranging from 20 mM to 100 mM. All solutions were composed of NaCl and Na_2HPO_4/NaH_2PO_4, buffered to pH 6.8.

Solution NMR assignments. Proton assignments were performed following well-established procedures; and in some cases aided by inosine substitutions. Identification of intranucleotide anomeric signals were derived from DQF-COSY and TOCSY experiments. G-quadruplex sequence specific assignments were based on sequential NOE connectivities of the type H8/H6(i)-H1'(i)-H8/H6(i+1) as well as the corresponding sequential connectivities on H2'/H2'' and H3' spin systems derived from NOESY experiments. JR NOESY experiments allowed for GH1 and GH21/GH22 assignments. Typically, (^{1}H-^{31}P)HSQC experiments that allow for tracing intranucleotide spin system connectivities involving H3'(i-1)-P(i)-H4'/H5'/H5''(i) were acquired in order to support or identify residue sequential connectivities. In select cases, unambiguous assignment of imino H1 from the aromatic H8 Guanines in the stem was also performed from natural abundance JR [^{1}H-^{13}C] HMBC experiments. The chemical shift assignments are shown in Tables S1-7**.**

Restraints for structure calculations. Distance restraints were typically derived from NOESY experiments in in "100%" ^2H_2O at 3 - 5 mixing times, and distances estimated from the initial buildup rates of the NOE curves by the two-spin approximation $r_{ij} = r_{ref}(R_{ij}/R_{ref})^{1/6}$, where r_{ij} is the distance between protons i and j, r_{ref} is a reference distance, and R_{ij} and R_{ref} are the initial buildup rates. Inter-proton distances were estimated using the average of the volume integral of the distance between H5–methyl in thymine residues, i.e. 2.46 Å. Only two limiting mixing times (60 and 200 ms) were used to derive distance restraints from the exchangeable protons collected with jump-and-return NOESY spectra at 5 °C in 90% H_2O, 10% ^2H_2O. Distances were assumed to be 3.0±0.8 Å for strong peaks observed in the 60 ms mixing time spectrum, 4.0±1.2 Å for medium cross-peaks observed in the 200 ms mixing time spectrum, and 5.0±1.8 Å for cross-peaks not observed in a 60 ms mixing time spectrum.

Structure calculations. Distance-restrained structure determinations were carried out using distance constraints from NMR data. Calculations were performed utilizing X-PLOR-NIH,(*41*) using the CHARMm force field and adapted for restrained molecular dynamics (rMD) for nucleic acids. All calculations were executed *in* v*acuo* without explicit counterions. Typically the distance geometry and simulated annealing refinement protocol started from 3000 different structures randomized over all dihedral angles. A number of structures did usually emerge with the same fold and separated from nonconverged structures by large gaps in components of the potential energy function (dihedral angles, van der Waals, NOE violations, and covalent geometry). These sets were subsequently submitted to rMD calculations, performed using random velocities fitting a Maxwell-Boltzmann distribution. The empirical energy function was developed for nucleic acids and treated all hydrogens explicitly. It

consisted of energy terms for hydrogen bonding, and nonbonded interactions, bonds, bond angles, torsion angles, and tetrahedral and planar geometry, including van der Waals and electrostatic forces. During these computations both the glycosidic torion angle c and planarity restraints were imposed during computations. The final procedure consisted of a total of 53 ps of rMD, including 7 ps of 14 ps from 300 to 1000 K, a 20 ps scaleup of restraints at high temperature, 14 ps of cooling to 300 K, and 12 ps of equilibration rMD, without planarity restraints. The temperature was controlled by coupling the molecules to a temperature bath with a coupling constant of 0.025 ps. The van der Waals term was approximated using the Lennard-Jones potential energy function, and bond lengths involving hydrogens were fixed with the SHAKE algorithm. NOE and dihedral angle restraints, chemical shifts, were deposited in the BMRB, and structure coordinates have been deposited in the Protein Data Bank (PDB) with identification codes 5J4P, 5J05, 5J4W, 5J6U, 2M6W, 2M6V, and 2MFT, as well as in the Biological Magnetic Resonance Bank (BMRB) with identification codes 30055, 30045, 30056, 30058, 19159, 19158, and 19571; respectively. Structural restraints are shown in Tables S8-14.

REFERENCES AND NOTES

1. D. Rhodes, H. J. Lipps, G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Research* **43**, 8627 (2015-Oct-15, 2015).
2. S. Neidle, Quadruplex Nucleic Acids as Novel Therapeutic Targets. *Journal Of Medicinal Chemistry* **59**, 5987 (Jul, 2016).
3. P. Alberti, J. L. Mergny, DNA duplex-quadruplex exchange as the basis for a nanomolecular machine. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **100**, 1569 (Feb 18, 2003).
4. Y. Sannohe, M. Endo, Y. Katsuda, K. Hidaka, H. Sugiyama, Visualization of Dynamic Conformational Switching of the G-Quadruplex in a DNA Nanostructure. *Journal Of The American Chemical Society* **132**, 16311 (Nov 24, 2010).

5. D. Koirala *et al.*, A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. *Nature Chemistry* **3**, 782 (Oct, 2011).
6. G. I. Livshits *et al.*, Long-range charge transport in single G-quadruplex DNA molecules. *Nature Nanotechnology* **9**, 1040 (Dec, 2014).
7. P. Travascio, Y. F. Li, D. Sen, DNA-enhanced peroxidase activity of a DNA aptamer-hemin complex. *Chemistry & Biology* **5**, 505 (Sep, 1998).
8. P. J. Bates, D. A. Laber, D. M. Miller, S. D. Thomas, J. O. Trent, Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Experimental and Molecular Pathology* **86**, 151 (Jun, 2009).
9. M. Webba da Silva, Geometric formalism for DNA quadruplex folding. *Chemistry European Journal* **13**, 9738 (2007).
10. A. I. Karsisiotis, C. O'Kane, M. Webba da Silva, DNA quadruplex folding formalism - A tutorial on quadruplex topologies. *Methods* **34**, 8 (2013 Nov, 2013).
11. M. Webba da Silva *et al.*, Design of a G-quadruplex topology through glycosidic bond angles. *Angew Chem Int Ed Engl* **48**, 9167 (2009).
12. C. J. Lech, A. T. Phan, M. E. Michel-Beyerle, A. A. Voityuk, Influence of Base Stacking Geometry on the Nature of Excited States in G-Quadruplexes: A Time-Dependent DFT Study. *Journal Of Physical Chemistry B* **119**, 3697 (Mar, 2015).
13. A. Randazzo, G. P. Spada, M. Webba da Silva, in *Quadruplex Nucleic Acids,* J. B. Chaires, D. Graves, Eds. (2013), vol. 330, pp. 67-86.
14. A. I. Karsisiotis *et al.*, Topological Characterization of Nucleic Acid G-Quadruplexes by UV Absorption and Circular Dichroism. *Angewandte Chemie-International Edition* **50**, 10645 (2011).
15. A. Guedin, J. Gros, P. Alberti, J. L. Mergny, How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic acids research* **38**, 7858 (2010).
16. A. Guedin, P. Alberti, J. L. Mergny, Stability of intramolecular quadruplexes: sequence effects in the central loop. *Nucleic Acids Research* **37**, 5559 (2009).
17. F. Fogolari *et al.*, Molecular models for intrastrand DNA G-quadruplexes. *Bmc Structural Biology* **9**, (Oct 7, 2009).
18. K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, Structure of the human telomere in K+ solution: An intramolecular (3+1) G-quadruplex scaffold. *Journal Of The American Chemical Society* **128**, 9963 (Aug 2, 2006).
19. A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic Acids Research* **35**, 6517 (2007).
20. M. Marusic, P. Sket, L. Bauer, V. Viglasky, J. Plavec, Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic Acids Research* **40**, 6946 (Aug, 2012).
21. K. W. Lim, V. C. M. Ng, N. Martin-Pintado, B. Heddi, A. T. Phan, Structure of the human telomere in Na+ solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Research* **41**, 10556 (Dec, 2013).
22. P. Galer, B. F. Wang, P. Sket, J. Plavec, Reversible pH Switch of Two-Quartet G-Quadruplexes Formed by Human Telomere. *Angewandte Chemie-International Edition* **55**, 1993 (Feb, 2016).

23. F. W. Smith, P. Schultze, J. Feigon, Solution structures of unimolecular quadruplexes formed by oligonucleotides containing Oxytricha telomere repeats. *Structure* **3**, 997 (Oct 15, 1995).
24. K. W. Lim *et al.*, Structure of the Human Telomere in K+ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *Journal Of The American Chemical Society* **131**, 4301 (Apr 1, 2009).
25. J. Brcic, J. Plavec, Solution structure of a DNA quadruplex containing ALS and FTD related GGGGCC repeat stabilized by 8-bromodeoxyguanosine substitution. *Nucleic Acids Research* **43**, 8590 (Sep, 2015).
26. L. Hu, K. W. Lim, S. Bouaziz, A. T. Phan, Giardia Telomeric Sequence d(TAGGG)(4) Forms Two Intramolecular G-Quadruplexes in K+ Solution: Effect of Loop Length and Sequence on the Folding Topology. *Journal Of The American Chemical Society* **131**, 16824 (Nov 25, 2009).
27. S. Amrane *et al.*, A novel chair-type G-quadruplex formed by a Bombyx mori telomeric sequence. *Nucleic Acids Research* **37**, 931 (Feb, 2009).
28. X. Cang, J. Sponer, T. E. Cheatham, III, Insight into G-DNA Structural Polymorphism and Folding from Sequence and Loop Connectivity through Free Energy Analysis. *Journal of the American Chemical Society* **133**, 14270 (Sep 14, 2011).
29. Y. Wang, D. J. Patel, Solution Structure of the Human Telomeric Repeat D Ag(3)(T(2)Ag(3))3 G-Tetraplex. *Structure* **1**, 263 (Dec 15, 1993).
30. X. H. Cang, J. Sponer, T. E. Cheatham, Explaining the varied glycosidic conformational, G-tract length and sequence preferences for anti-parallel G-quadruplexes. *Nucleic Acids Research* **39**, 4499 (May, 2011).
31. A. I. Karsisiotis, M. Webba da Silva, Structural Probes in Quadruplex Nucleic Acid Structure Determination by NMR. *Molecules* **17**, 13073 (Nov, 2012).
32. A. Matsugami, Y. Xu, Y. Noguchi, H. Sugiyama, M. Katahira, Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a K+ solution. *Febs Journal* **274**, 3545 (Jul, 2007).
33. Y. Xu, Y. Noguchi, H. Sugiyama, The new models of the human telomere d[AGGG(TTAGGG)(3)] in K+ solution. *Bioorganic & Medicinal Chemistry* **14**, 5584 (2006).
34. C. F. Tang, R. H. Shafer, Engineering the quadruplex fold: Nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *Journal Of The American Chemical Society* **128**, 5966 (May 3, 2006).
35. J. Zhou *et al.*, Unexpected Position-Dependent Effects of Ribose G-Quartets in G-Quadruplexes. *Journal Of The American Chemical Society* **139**, 7768 (Jun, 2017).
36. M. Marusic, J. Plavec, The Effect of DNA Sequence Directionality on G-Quadruplex Folding. *Angewandte Chemie-International Edition* **54**, 11716 (Sep, 2015).
37. P. Hazel, J. Huppert, S. Balasubramanian, S. Neidle, Loop-length-dependent folding of G-quadruplexes. *Journal of the American Chemical Society* **126**, 16405 (Dec 22, 2004).
38. M. Liu *et al.*, Magnetically Actuated Wormlike Nanomotors for Controlled Cargo Release. *Acs Applied Materials & Interfaces* **7**, 26017 (Dec, 2015).

39. L. Lv, Z. J. Guo, J. H. Wang, E. K. Wang, G-Quadruplex as Signal Transducer for Biorecognition Events. *Current Pharmaceutical Design* **18**, 2076 (May, 2012).
40. N. M. Hessari *et al.*, Programmed Self- Assembly of a Quadruplex DNA Nanowire. *Chemistry-A European Journal* **20**, 3626 (Mar, 2014).
41. C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, The Xplor-NIH NMR molecular structure determination package. *Journal Of Magnetic Resonance* **160**, 65 (Jan, 2003).
42. H. M. Berman *et al.*, The Protein Data Bank. *Nucleic Acids Research* **28**, 235 (Jan 1, 2000).
43. Y. Wang, D. J. Patel, Solution structure of the Oxytricha telomeric repeat d[G4(T4G4)3] G-tetraplex. *J Mol Biol* **251**, 76 (Aug 4, 1995).
44. Z. J. Zhang, J. X. Dai, E. Veliath, R. A. Jones, D. Z. Yang, Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K+ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Research* **38**, 1009 (2010).

ACKNOWLEDGEMENTS

This research was supported by the BBSRC (grant number BB/H005692) to MWS, and a Vice-Chancellor Research Studentship (University of Ulster) to SAD. We acknowledge use of the open-access NMR facilities: (i) at HWB-NMR, and the associated Wellcome Trust Biomedical Resources (083796/Z/07/Z); (ii) through EU FP7 EAST-NMR (228461) at the Slovenian NMR Center.

AUTHOR CONTRIBUTIONS

MWS conceived the project; MWS, SAD developed concepts; MWS, SAD, AIK, designed primary sequences; AIK, SAD, MWS acquired data; SAD, AIK, MWS analyzed experimental data; SAD, MWS performed structure calculations; SAD, MWS, AIK drafted the manuscript.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FIGURES AND TABLES

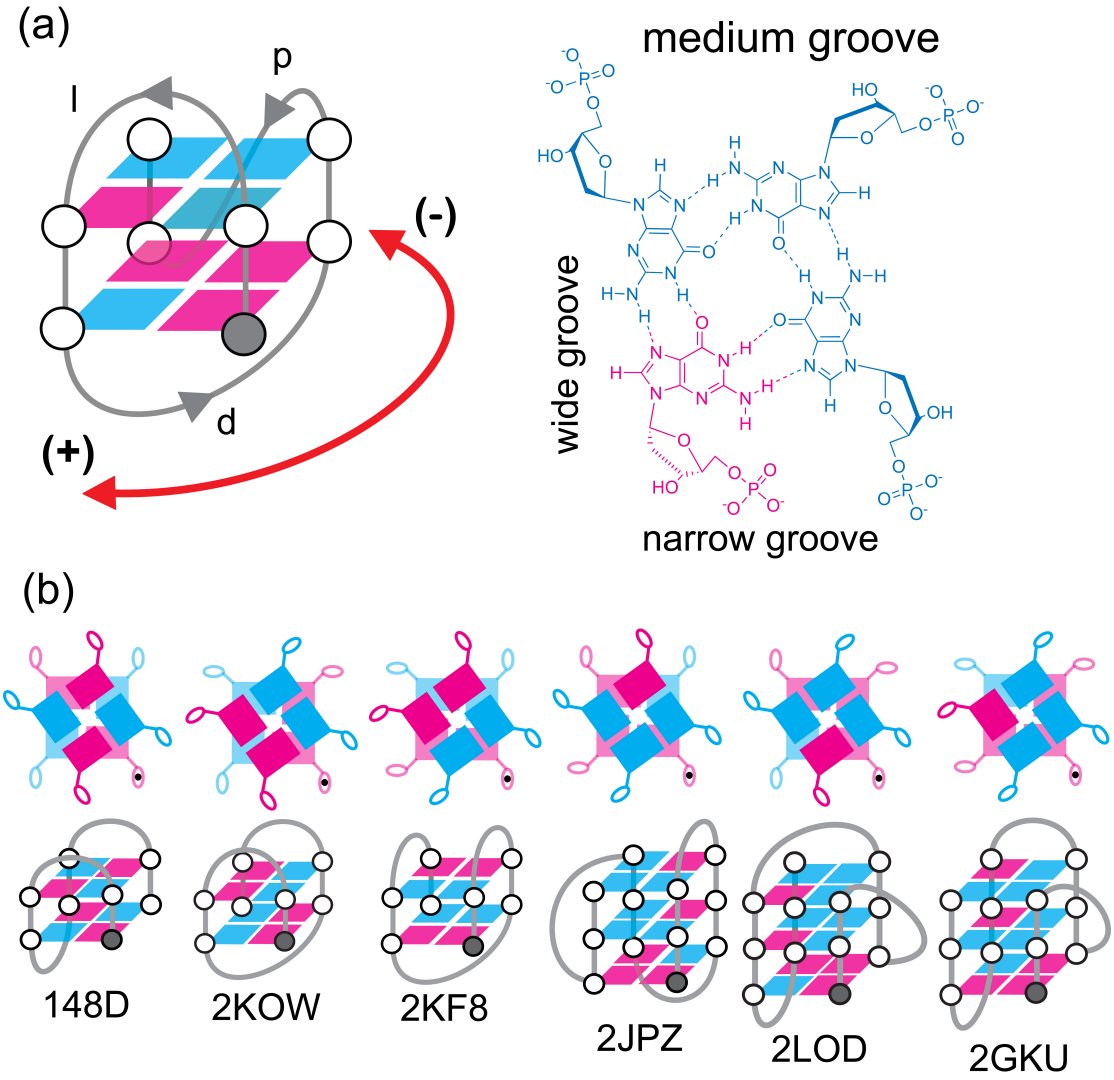


Fig. 1. Structural descriptors of canonical quadruplexes. In (A) the 2(+l_nd-p) folding topology and hydrogen-bond alignments for its top tetrad are shown. Magenta denotes *syn* glycosidic bond angles and cyan denotes *anti*. The grey circle indicates the 5'-end in the stem. Strand directionalities are indicated by (-) when counter-clockwise and (+) clockwise. Propeller loops are indicated by the symbol "p", diagonal loops by "d", and lateral loops by "l". In (B) schematic representations of named high-resolution solution structures publicly available in the Protein Data Bank (PDB)(*42*) are shown, with corresponding PDB ID and the respective schematic of the quadruplex topology. The topology of the two-stacked thrombin binding aptamer, "TBA" (PDB id 148D) is known as "chair-type" quadruplex. It

can be described as a quadruplex adopting the 2(+l_n+l_w+l_n) topology. The two two-stacked "basket-type" architectures of human telomeric repeats (2KF8) and Giardia telomeric repeats (2KOW) are denoted 2(-l_wd+l_n) and 2(+l_nd-l_w); respectively. The three-stacked "Form-1" and "Form-2" topologies of human telomeric sequences are described by 3(-p-l_n-l_w) for (2GKU,) and 3(-l_w-l_n-p) for (2JPZ,), respectively. Finally, the "(3+1) scaffold" of 2LOD can be named 3(-pd+l_n).

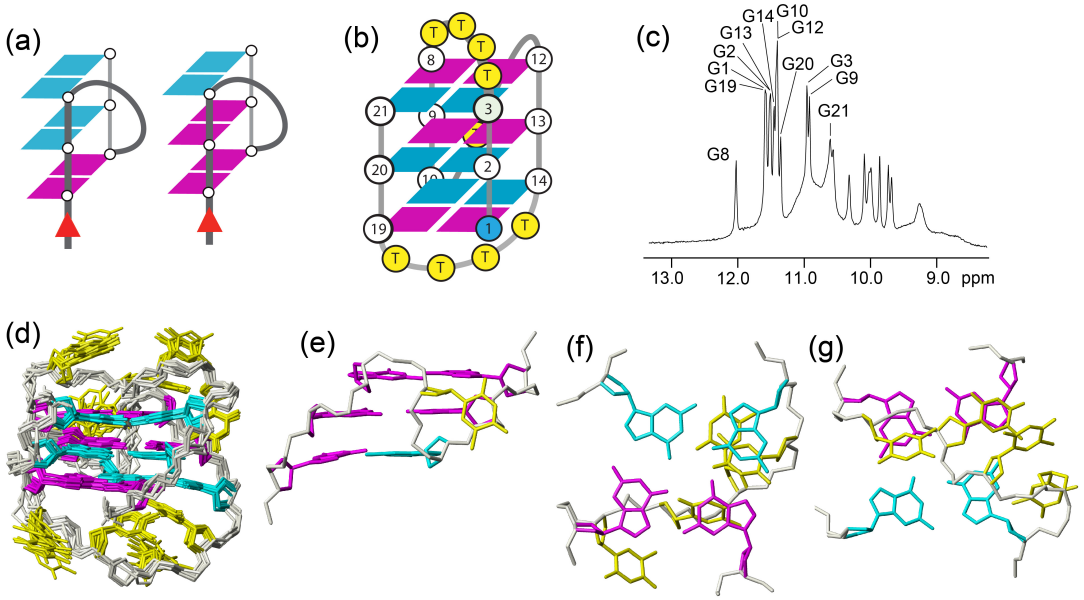


Fig. 2. Experimental verification of propeller loops bridging parallel stranded *syn*G-*syn*G-*anti*G grooves. In (a) schematic representations of grooves composed of parallel-stranded *syn*G-*anti*G-*anti*G (left) and *syn*G-*syn*G-*anti*G (right) segments. The 2'-deoxyguanosines are shown as *syn* (magenta) and *anti* (cyan) conformations. In (b) the design of the 3(d+pd) topology using the DNA sequence d($G_3T_4G_3TG_3T_4G_3$); where a single thymine was used to form a propeller loop, and 4-thymine segments to form diagonal loops. In (c) expansion of the proton NMR spectrum at 5 ^oC illustrates the twelve assigned imino protons indicating formation of a 3-stacked quadruplex fold in 16 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8. In (d) family of ten superpositioned refined structures of the quadruplex formed by this sequence in solution. In (e), view into the medium grove populated by the single thymine of the propeller clockwise loop. In (f), bird's-eye view of the disposition of the second diagonal loop capping the stem over the (G1:G14:G10:G19) tetrad. In (g) bird's-eye view of disposition of the first diagonal loop over the (G3:G12:G8:G21) tetrad.

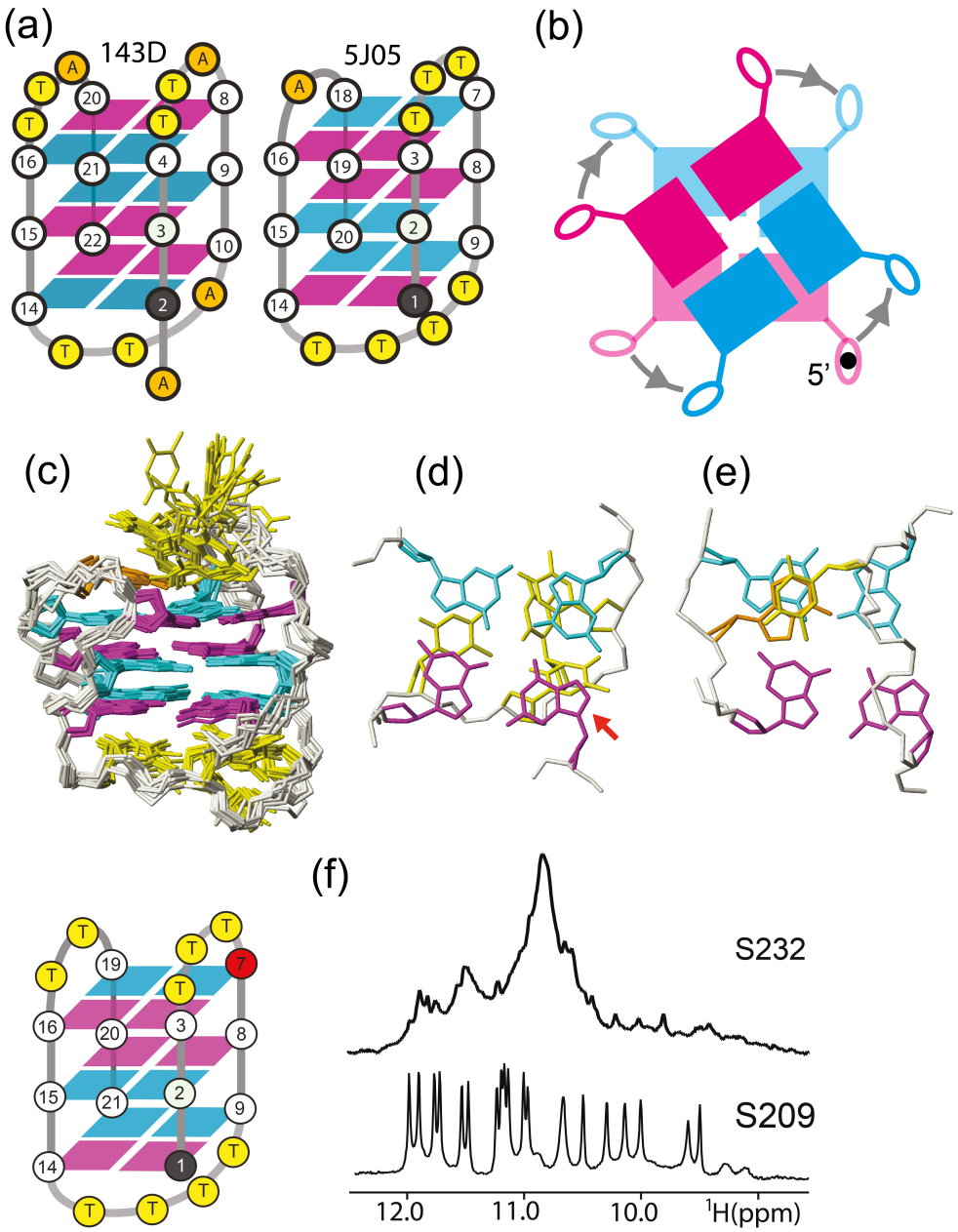


Fig. 3. Alternative conformations of glycosyl bonds in a quadruplex stem are possible. In (a) schematic representations of the alternative sequence of glycosidic bond angles in the 3($-l_w d+l_n$) topology for PDB id 143D (left) and the designed 5J05, and corresponding groove-width combinations (b) in the stem. In (c), a bundle of solution structures adopted by the DNA sequence 5J05 in 100 mM sodium solution at pH 6.8. In (d) the capping of the diagonal onto the (G1:G9:G20:G14) tetrad, with the arrow indicating the position of *syn*G1. In (e) detail of the intrastrand stacking of the interdigitated adenine stacking onto *anti*G18 of the stem and T6. In (f) design of the 3($-l_w d+l_n$) topology by replacing the nucleoside dG7 of the DNA sequence S232 by a rG in S209, as shown in red in the schematics for the topology. In the expansion of the proton NMR spectrum in 0.1 M sodium solutions at 5 ^oC shown, the imino protons in the spectrum of S232 appear predominantly as 'hump' at approximately 10.8 ppm indicating that the sequence is mostly unfolded. Nineteen of the 21 possible imino protons in the DNA sequence appear in the spectrum of S209 indicating formation of a 3-stacked quadruplex fold. The topology has been structurally characterized (Supplementary Material).

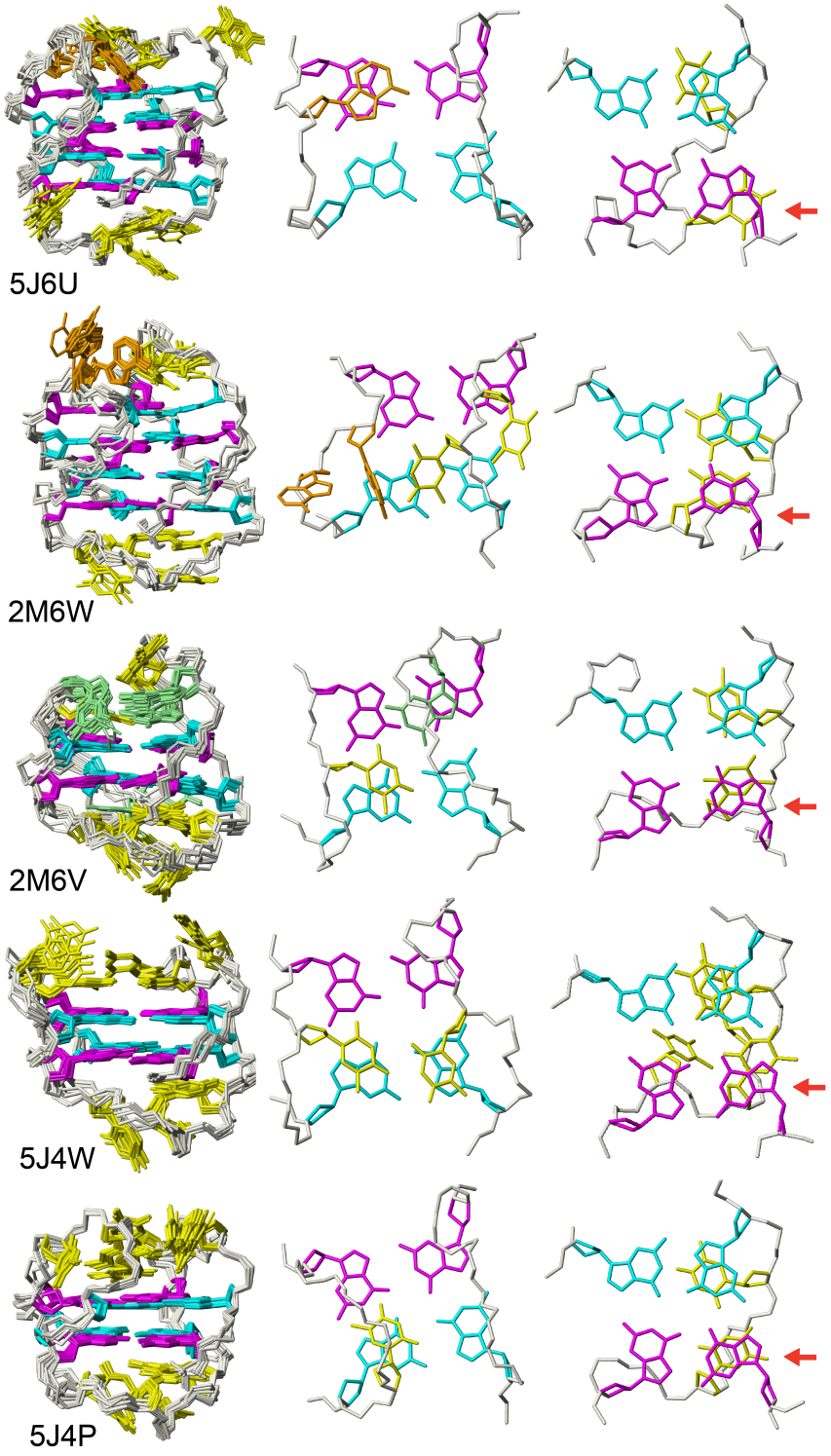


Fig. 4. Structural details for high-resolution structures designed to adopt the ($-l_w d+l_n$) topology. The DNA sequences 5J6U and 2M6W adopt 4($-l_w d+l_n$), 2M6V, 5J4W, and 5J4P adopt 3($-l_w d+l_n$) in 100 mM sodium solution at pH 6.8. For each structure details of loop stacking interactions with the top (left most) and bottom (right) tetrads are depicted. The red arrows indicate proximity of the 5'-*syn*G residue of the stem to the third thymine in the diagonal loop. Shown are also 2'-deoxyguanosines of the stem in *syn* (magenta) and *anti* (cyan) conformations, green for non-stem guanosines, and yellow for thymines. In all structures the first thymine of the diagonal loop stacks onto the preceding guanosine of the stem.

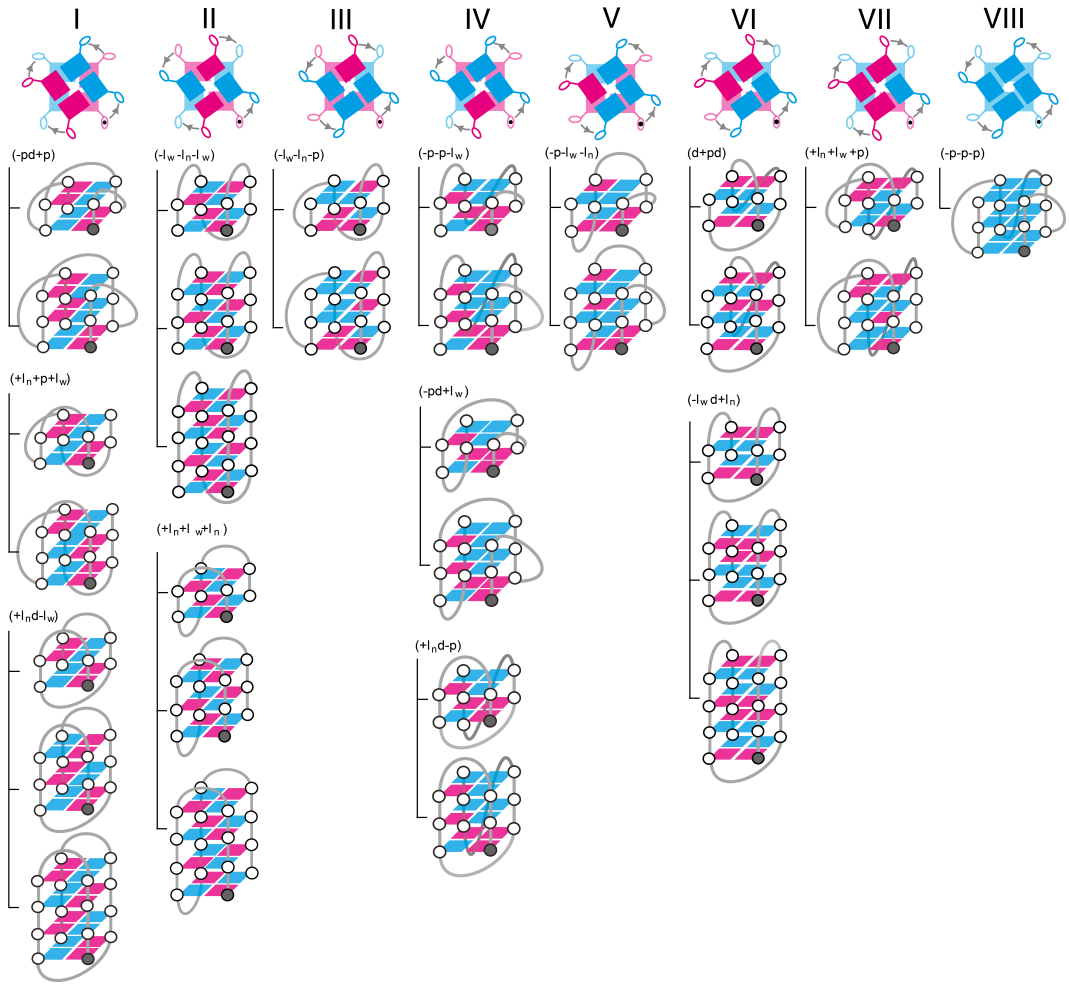


Fig. 5. Schematic description of all canonical quadruplex topologies feasible with their constituent groove type combination. In the top, *syn* (magenta) and *anti* (cyan) guanosines combined through hydrogen bond alignments to form tetrads representing the eight possible quadruplex stems. Indicated in grey are the only possible strand progression directionalities for each stem.

Table 1. Oligonucleotide sequences designed to study folding of the ($-l_w d+l_n$) topology. Guanosines in *syn* conformation appear in bold when determined.

Name	PRE	G1	L1	G2	L2	G3	L3	G4	POS	Status*
S036	-	**GGG**G	TT	**GGG**G	TTTT	**GGG**G	T	**GGG**G	-	Top
S141	-	GGGG	TT	GGGG	TTTT	GGGG	TT	GGGG	-	Ms
2M6W	-	**GGG**G	TT	**GGG**G	TTTT	**GGG**G	AA	**GGG**G	-	Str
S069	-	**GGG**G	TT	**GGG**G	TTTT	**GGG**G	TTT	**GGG**G	-	Top
S067	-	**GGG**G	AA	**GGG**G	TTTT	**GGG**G	TTT	**GGG**G	-	Top
S064	-	GGGG	TTT	GGGG	TTTT	GGGG	A	GGGG	-	Ms
5J6U	-	**GGG**G	TTT	**GGG**G	TTTT	**GGG**G	AA	**GGG**G	-	Str
S080	-	**GGG**G	TTT	**GGG**G	TTTT	**GGG**G	TTT	**GGG**G	-	Top
S066	-	GGGG	TTTT	GGGG	TTTT	GGGG	AA	GGGG	-	Ms
201D	-	**GGG**G	TTTT	**GGG**G	TTTT	**GGG**G	TTTT	**GGG**G	-	(*43*)
230D	-	**GGG**G	TUTU	**GGG**G	TTTT	**GGG**G	UUTT	**GGG**I	-	(*23*)
S025	-	GGG	TT	GGG	TTTT	GGG	A	GGG	-	Ms
S087	-	GGG	TT	GGG	TTTT	GGG	AA	GGG	-	Ms
S089	-	**GGG**	TT	G**G**G	TTTT	**GGG**	TTT	G**G**G	-	Top
S088	-	**GGG**	AA	G**G**G	TTTT	**GGG**	TTT	G**G**G	-	Top
5J05	-	**GGG**	TTT	G**G**G	TTTT	**GGG**	A	G**G**G	-	Str
S029	-	GGG	TTT	GGG	TTTT	GGG	T	GGG	-	Ms
S093	T	**GGG**	TTT	G**G**G	TTTT	**GGG**	A	G**G**G	-	Top
S090	-	**GGG**	TTT	G**G**G	TTTT	**GGG**	AA	G**G**G	-	Top
S232	-	GGG	TTT	GGG	TTTT	GGG	TT	GGG	-	Ms
S209	-	**GGG**	TTT	(rG)**G**G	TTTT	**GGG**	TT	G**G**G	-	Top
S210	-	GGG	TTT	GGG	TTTT	(rG)GG	TT	GGG	-	Ms
S231	-	**GGG**	TTT	G**G**G	TTTT	**GGG**	TTT	G**G**G	-	Top
S174	-	GGG	TTTT	GGG	TTTT	GGG	TT	GGG	-	Ms
S175	-	GGG	TTTT	GGG	TTTT	GGG	TTT	GGG	-	Ms
S038a	-	GGG	TTTT	GGG	TTTT	GGG	TTTT	GGG	-	Ms
S166	-	GG	TT	GG	TTTT	GG	T	GG	-	Ms
S169	-	GG	TT	GG	TTTT	GG	TT	GG	-	**
S230	A	GG	TT	GG	TTTT	GG	TT	GG	-	Ms
S172	-	**G**G	TT	**G**G	TTTT	**G**G	TTT	**G**G	-	Top
S167	-	**G**G	TTT	**G**G	TTTT	**G**G	T	**G**G	-	Top
5J4W	-	**G**G	TTT	**G**G	TTTT	**G**G	TT	**G**G	-	Str
S205	A	GG	TTT	GG	TTTT	GG	TT	GG	-	Ms
5J4P	-	**G**G	TTT	**G**G	TTTT	**G**G	TTT	**G**G	-	Str
S168	-	GG	TTTT	GG	TTTT	GG	T	GG	-	Ms
S171	-	**G**G	TTTT	**G**G	TTTT	**G**G	TT	**G**G	-	Top
2M6V	-	**G**G	GTTG	**G**G	TTTT	**G**G	GT	**G**G	G	Str
2KF8	-	**G**G	GTTA	**G**G	GTTAG	**G**G	TTA	**G**G	GT	(*24*)
2KKA	A	**G**G	GTTA	**G**G	GTTAI	**G**G	TTA	**G**G	GT	(*44*)

*Ms stands for 'multiple species', 'Str' for high-resolution structure determined, and 'Top' for characterization of ($-l_w d+l_n$) topology by solution NMR methods. ** Not ($-l_w d+l_n$). Except for 2KF8 (90 mM K^+ solution pH 7) and 2KKA (95 mM K^+ solution pH 7) folding was evaluated in 100 mM Na^+ solution pH 6.7.

SUPPLEMENTARY MATERIAL

fig. S1. Expansions of 1D NMR spectra of imino proton regions for DNA sequences folding into quadruplexes in this study.

table S14. NMR restraints for and structural statistics for structures of 2M6V.

SUPPLEMENTARY MATERIAL

ENCODING CANONICAL DNA QUADRUPLEX STRUCTURE

S. A. Dvorkin, A. I. Karsisiotis, M. Webba da Silva*

Biomedical Sciences Research Institute, Ulster University, Coleraine, BT52 1SA, UNITED KINGDOM.

* To whom correspondence should be addressed. Tel: 44 2870124009; Email: mm.webba-da-silva@ulster.ac.uk

SUPPLEMENTARY MATERIAL CONTENT

A. ASSESSMENT OF FOLDING OF DNA SEQUENCES (TABLE 1)

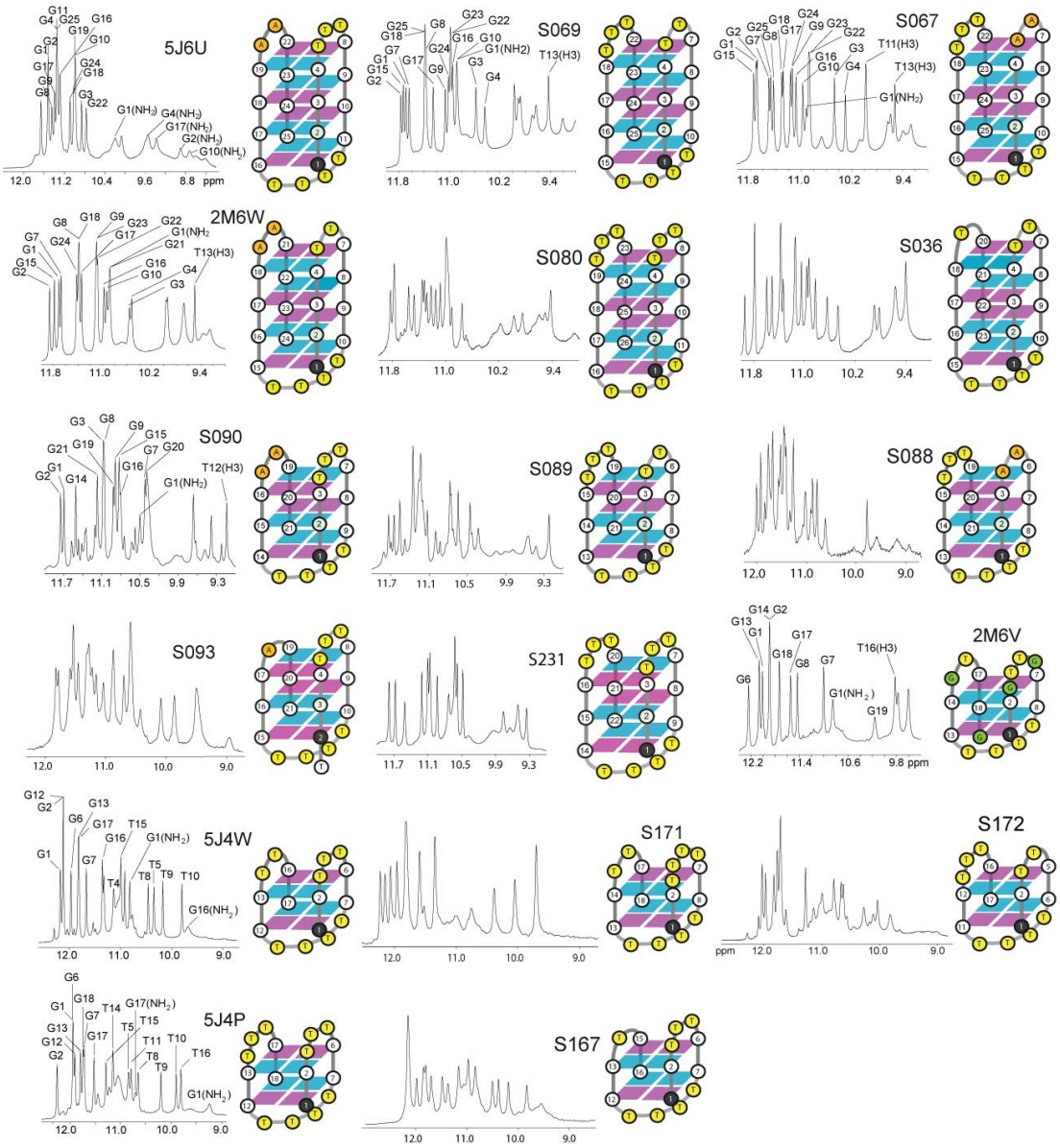


Figure S1. Expansions of 1D NMR spectra of imino proton regions for DNA sequences at ~2 mM oligonucleotide concentrations in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 solutions at at 5 ^oC. Shown are also schematic representations of the ($-l_wd+l_n$) topology they adopt with 2'-deoxyguanosines of the stem in *syn* (magenta) and *anti* (cyan) conformations, green for non-stem guanosines, and orange for adenines, and yellow for thymines.

B. CHARACTERIZATION OF STRUCTURE

Structural assignments for 2MFT in sodium

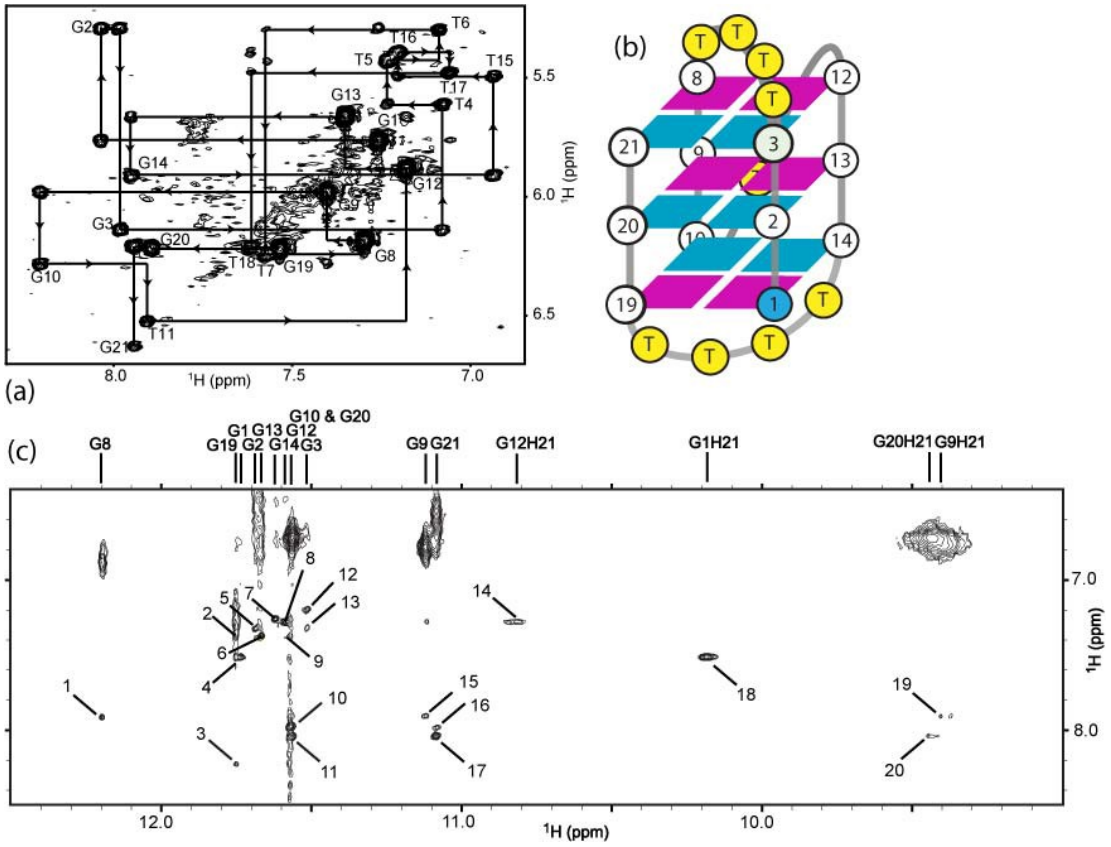


Figure S2. The DNA sequence d($G_3T_4G_3TG_3T_4G_3$) at 2 mM concentration adopts the 3(d+pd) topology in 16 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 (PDB id 2MFT). Anomeric-aromatic at 25 °C (A) and, aromatic-imino at 5 °C (B) regions of the ^{1}H,^{1}H NOESY spectrum of 2MFT are shown, along with a schematic for the topology designed for this DNA sequence (C). In panel (A), intraresidual H1′-H6/H8 NOE interactions are labelled and sequential correlations are denoted with lines. In panel (B), H8-H1 cross-peaks within G1-G14-G10-G19, G2-G13-G9-G20, and G3-G12-G8-G21 tetrads are labeled. Peaks 1-20 are assigned as follows: (1) G8H1-G21H8, (2) G19H1-G9H8, (3) G19H1-G10H8, (4) G1H1-G19H8, (5) G2H1-G13H8, (6) G13H1-G9H8, (7) G14H1-G1H8, (8) G12H1-G8H8, (9) G12H1-G9H8, (10) G10H1-G14H8, (11) G20H1-G2H8, (12) G3H1-G12H8, (13) G3H1-G13H8, (14) G12H21-G8H8, (15) G9H1-G20H8, (16) G21H1-G3H8, (17) G21H1-G2H8, (18) G1H21-G19H8, (19) G9H21-G20H8 and (20) G20H21-G2H8. A complete list of proton assignments is shown in **Table S1**.

Structural assignments for 2M6W in sodium

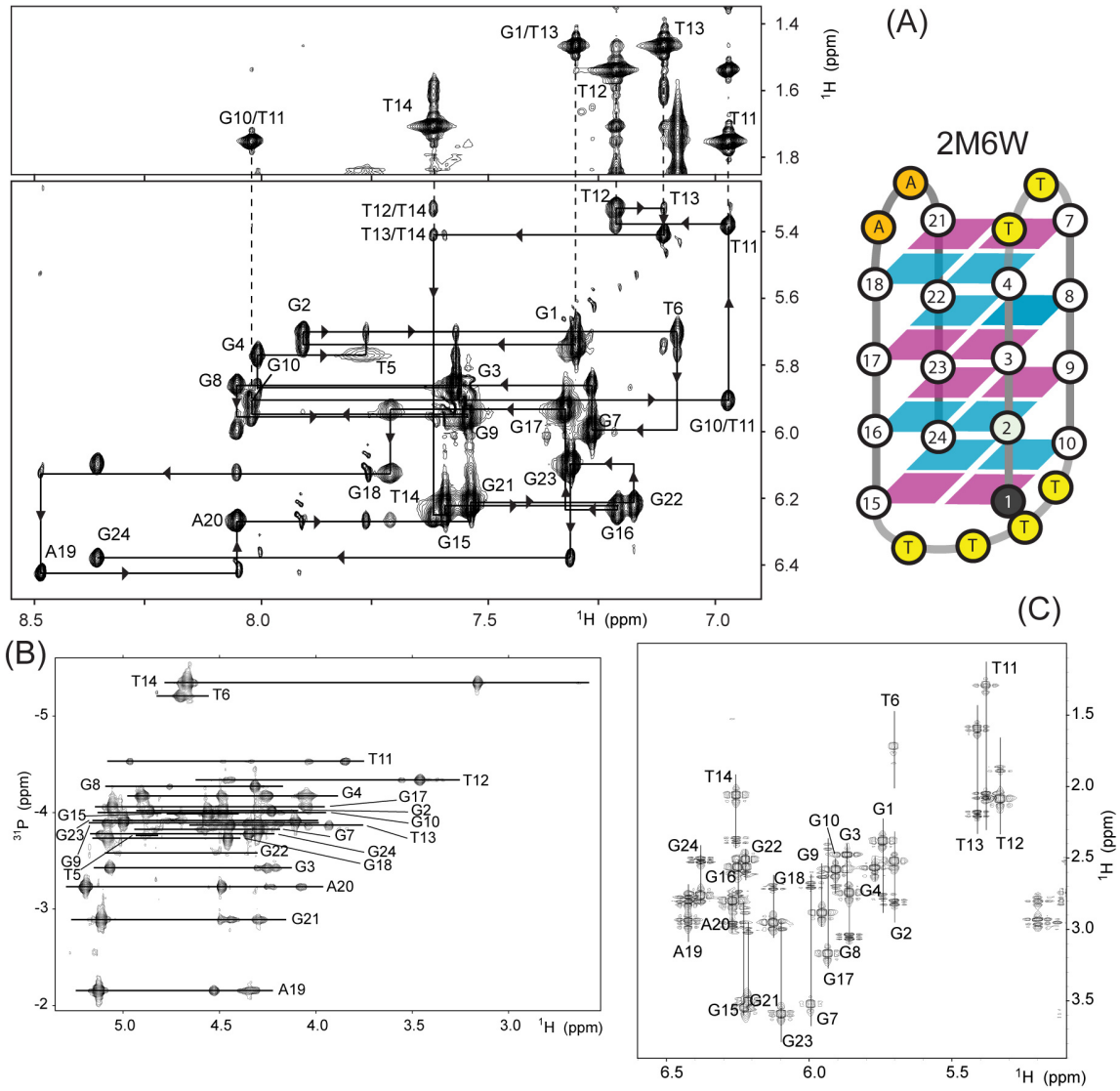


Figure S3. Non-exchangeable ^{1}H and ^{31}P assignments for the 4(-l_wd+l_n) adopted by the DNA sequence 2M6W in 16 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8, at 20 °C. In panel (A), expansions of ^{1}H-^{1}H NOESY spectra (20 ^{o}C) depicting anomeric-aromatic regions of the ^{1}H-^{1}H NOESY and showing labelled intraresidual H1′-H6/H8 and H6/H8-H2’/2’’ NOE interactions. Sequential correlations are denoted with lines. The inset contains assignments for the characteristic sequential connectivities (*Syn*G-*Anti*G-T-T-T-T) of the diagonal loop. Methyl-H8/H6 sections illustrate the characteristic connectivity between the aromatic H8 of the 5’-*Syn*G residue of the stem and the methyl of the third Thymine in the diagonal loop. Shown are also schematic representations of the topologies they adopt with 2’-deoxyguanosines of the stem in *syn* (pink) and *anti* (cyan) conformations, green for non-stem guanosines, and yellow for thymines. In panel (B) sequential coupling correlations of the type H3’(i-1)-P(i)-H4’/H5’/H5’’ in a[^{1}H-^{31}P] HSQC spectrum are shown. In panel (C) DQF COSY intraresidue correlations H1’-H2’/H2’’ are shown. A complete list of proton assignments is shown in **Table S2**.

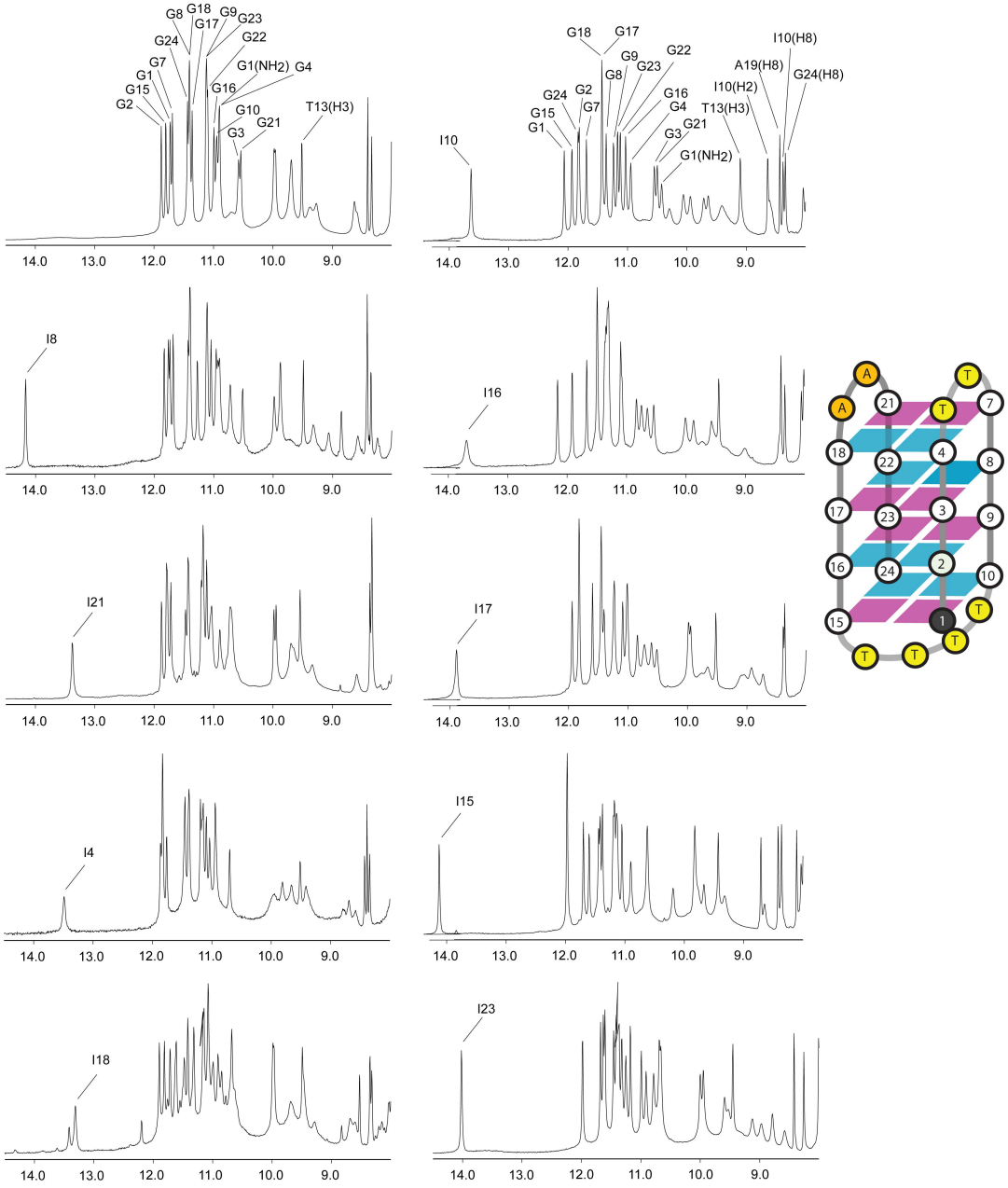


Figure S4. Exchangeable proton region expansions of NMR spectra of DNA sequence 2M6W (top left) and of its inosine substitutions (indicated) in sodium solutions at pH 6.8.

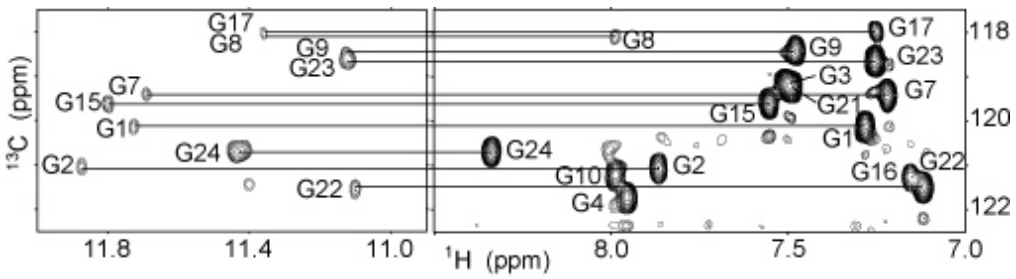


Figure S5. Heteronuclear natural abundance JR [^{1}H-^{13}C] H1(C5)H8 HMBC spectrum of 2M6W in 16 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8. Expansion illustrates the long range imino H1 to

aromatic H8 coupling through the $^{13}C_5$ atom, in $^{1}H_2O$ at 15 °C, providing unambiguous imino proton assignments for Guanines in the stem.

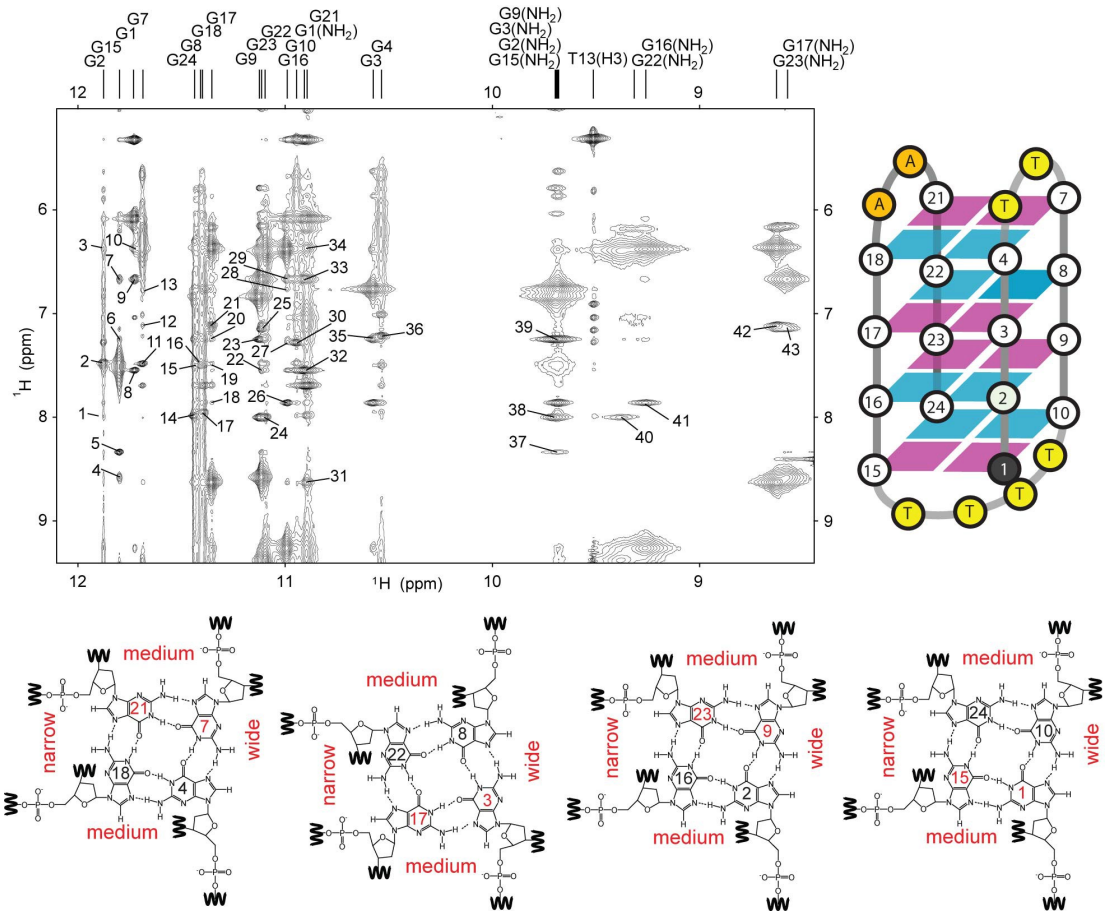


Figure S6. Exchangeable proton assignments for sequence 2M6W in 16 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 at 5° C. An expansion of a JR-NOESY spectrum (250 ms) at $^{2}H_2O$ at 5 °C is shown, illustrating the dipolar connectivities between imino and amino exchangeable protons with aromatic protons. Peaks 1-43 are assigned as follows: (1) G2H1-G10H8, (2) G2H1-G9H8, (3) G2H1-G16H22, (4) G15H1-G23H21, (5) G15H1-G24H8, (6) G15H1-G23H8, (7) G15H1-G23H22, (8) G1H1-G15H8, (9) G1H1-G23H22, (10) G7H1-G21H8, (11) G7H1-G22H8, (12) G7H1-G3H22, (13) G24H1-G10H8, (14) G24H1-G15H22, (15) G8H1-G3H8, (16) G18H1-G4H8, (17) G17H1-G2H8, (18) G17H1-G21H8, (19) G17H1-G23H8, (20) G17H1-G22H8, (21) G23H1-G15H8, (22) G9H1-G23H8, (23) G22H1-G8H8, (24) G23H1-G16H8, (25) G16H1-G2H8, (26) G16H1-G1H8, (27) G16H1-G3H22, (28) G16H1-G23H22, (29) G10H1-G1H8, (30) G21H1-G17H21, (31) G1H21-G15H8, (32) G1H21-G23H22, (33) G21H1-G17H22, (34) G3H1-G17H8, (35) G4H1-G7H8, (36) G15H21-G24H8, (37) G2H21-G8H8, (38) G3H21-G17H8, (39) G22H21-G8H8, (40) G16H21-G2H8, (41) G17H21-G22H8, (43) G23H21-G16H8. These assignments allow for the formation of the hydrogen bond alignments shown in the chemical structures below defining the topology shown.

Structural assignments for 5J6U in sodium

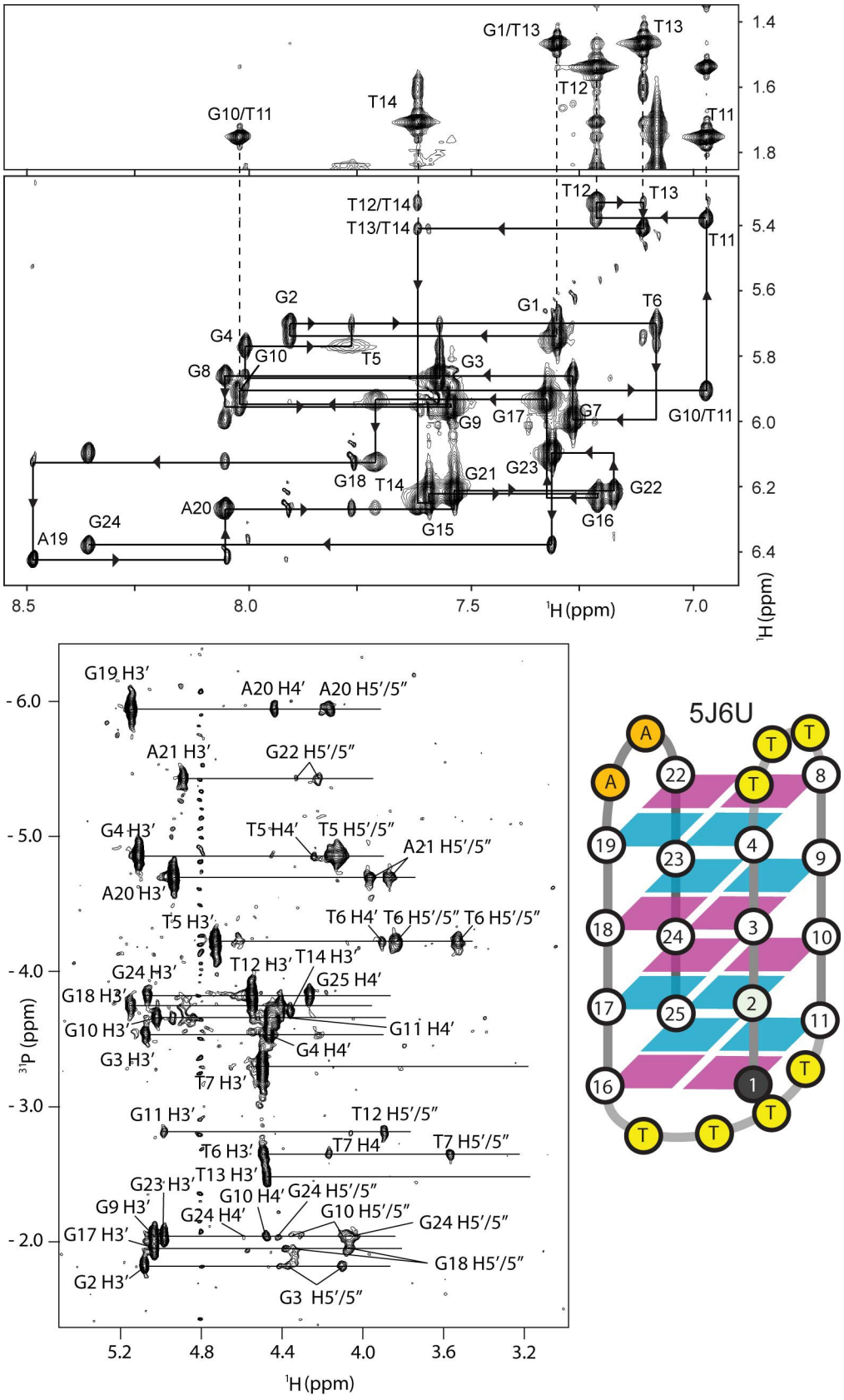


Figure S7. Non-exchangeable ^{1}H and ^{31}P assignments for the 4(-l_wd+l_n) adopted by the DNA sequence 5J6U in 16 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8, at 20 °C. The spectrum on top shows expansions of ^{1}H-^{1}H NOESY spectra (20 °C) depicting anomeric-aromatic regions of the ^{1}H-^{1}H

NOESY and showing labelled intraresidual H1′-H6/H8 and H6/H8-H2'/2'' NOE interactions. Sequential correlations are denoted with lines. The inset contains assignments for the characteristic sequential connectivities (*Syn*G-*Anti*G-T-T-T-T) of the diagonal loop. Methyl-H8/H6 sections illustrate the characteristic connectivity between the aromatic H8 of the 5'-*Syn*G residue of the stem and the methyl of the third Thymine in the diagonal loop. Shown are also schematic representations of the topologies they adopt with 2'-deoxyguanosines of the stem in *syn* (pink) and *anti* (cyan) conformations, orange for adenines, and yellow for thymines. In the bottom spectrum sequential coupling correlations of the type H3'(i-1)-P(i)-H4'/H5'/H5'' in a [^{1}H-^{31}P] HSQC spectrum are shown. A complete list of proton assignments is shown in **Table S3**.

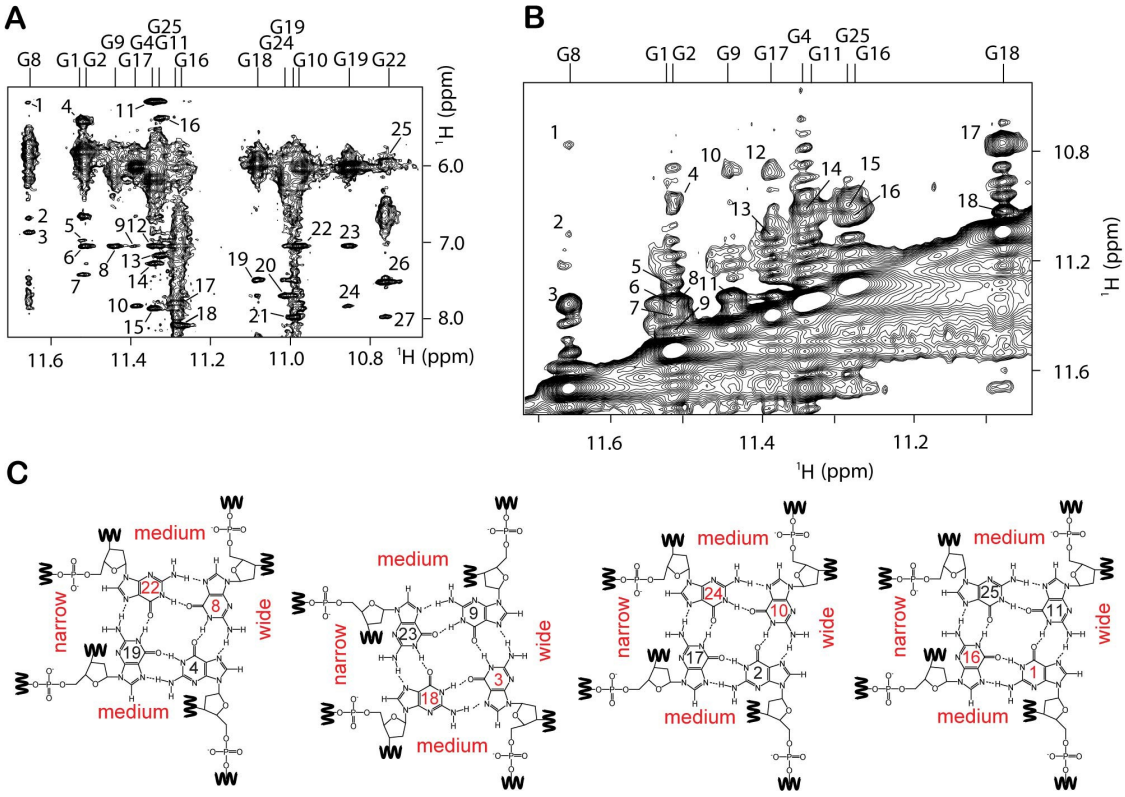


Figure S8. Exchangeable proton assignments for 5J6U. In (A) an expansion of a JR-NOESY spectrum (240 ms) in 16 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 at 5° C is shown, illustrating the dipolar connectivities between imino and amino exchangeable protons with aromatic protons. Peaks 1-27 are assigned as follows: (1) G8H1-T6H1', (2) G8H1-T7H8, (3) G8H1-G22H8, (4) G1H1-T14H1', (5)G1H1-T14H6, (6) G2H1-G10H8, (7) G1H1-G16H1, (8) G9H1-G3H8, (9) G17H1-G3H8, (10) G17H1-G2H8, (11) G4H1-T6H1', (12) G4H1-G3H1, (13) G11H1-G1H1, (14) G4H1-G8H8, (15) G4H1-G9H8, (16) G11H1-T12H1', (17) G25H1-G11H8, (18) G16H1-G25H1, (19) G18H1-G23H8, (20) G24H1-G17H8, (21) G19H1-G4H8, (22) G10H1-G24H8, (23) G3H1-G18H8, (24) G3H1-G2H8, (25) G22H1-A21H1', (26) G22H1-A21H1', (27) G22H1-G19H8. In (B) exchangeable proton assignments for 4[T3,A2]. An expanded JR-NOESY spectrum (200 ms) in 16 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 at 5 °C is shown, illustrating the dipolar connectivities between imino H1-H1 exchangeable protons. Peaks 1-18 are assigned as follows: (1) G8H1-G22H1, (2) G8H1-G18H1, (3) G8H1-G4H1, (4) G2H1-G10H1, (5) G1H1-G16H1, (6) G1H1-G11H1, (7) G1H1-G17H1, (8) G2H1-G11H1, (9) G2H1-G9H1, (10) G9H1-G3H1, (11) G9H1-G4H1, (12) G17H1-G3H1, (13) G17H1-G18H1, (14) G4H1-G19H1, (15) G25H1-G10H1, (16) G16H1-G24H1, (17) G18H1-G22H1, (18) G18H1-G24H1. These assignments allow for the formation of the hydrogen bond alignments shown in the chemical structures shown in (C).

Structural assignments for 5J05 in sodium

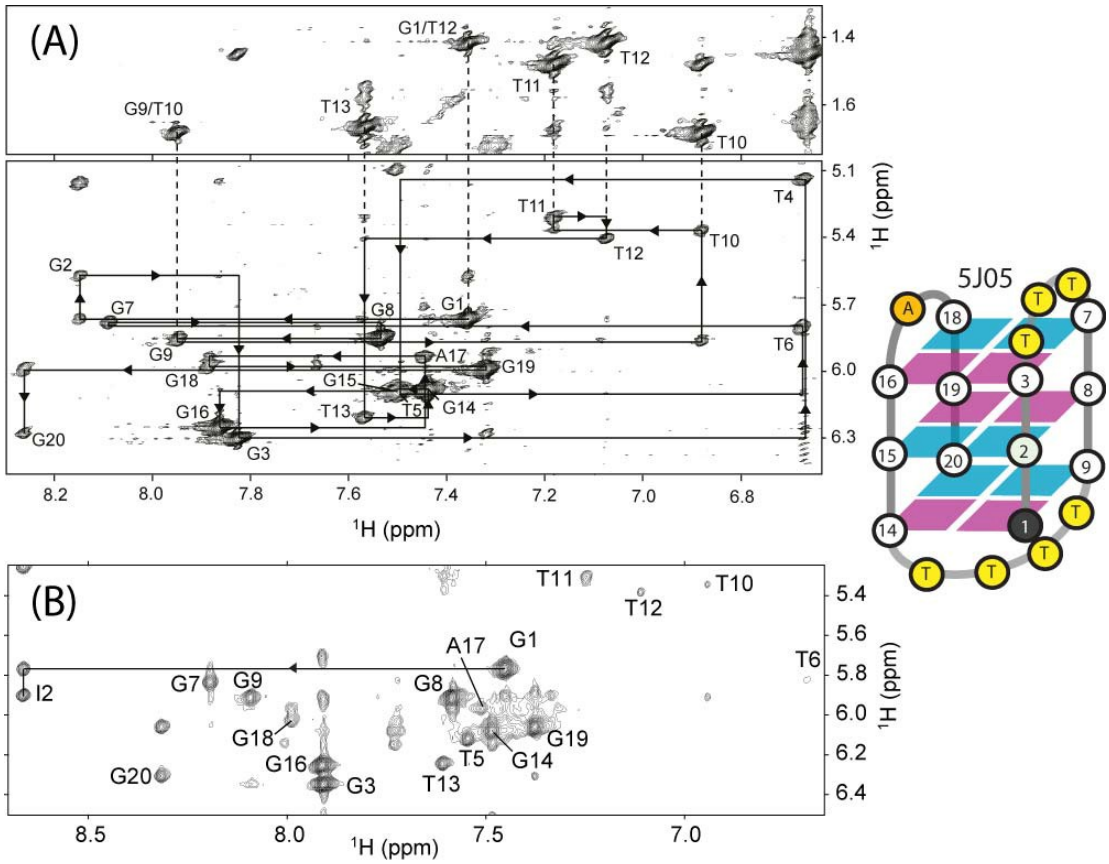


Figure S9. Non-exchangeable ^{1}H assignments for the 3(-l_wd+l_n) adopted by the DNA sequence 5J05 in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8. In panel (A), expansions of ^{1}H-^{1}H NOESY spectra (20 ^{o}C) depicting anomeric-aromatic regions of the ^{1}H-^{1}H NOESY and showing labelled intraresidual H1′-H6/H8 and H6/H8-H2'/2'' NOE interactions. Sequential correlations are denoted with lines. The inset contains assignments for the characteristic sequential connectivities (*Syn*G-*Anti*G-T-T-T-T) of the diagonal loop. Methyl-H8/H6 sections illustrate the characteristic connectivity between the aromatic H8 of the 5'-*Syn*G residue of the stem and the methyl of the third Thymine in the diagonal loop. Shown are also schematic representations of the topologies they adopt with 2'-deoxyguanosines of the stem in *syn* (pink) and *anti* (cyan) conformations, orange for adenines, and yellow for thymines. In panel (B) an expansion of JR [^{1}H-^{1}H] NOESY spectrum (200 ms mixing time) in $^{1}H_2O$ at 5 °C of the substitution of G2 for Inosine in DNA sequence 5J05. The spectrum illustrates the chemical shift of H8 upon substitution, which unambiguously proves the assignment of the sequentially linked G1 as a *syn* residue for 5J05. A complete list of proton assignments is shown in **Table S4**.

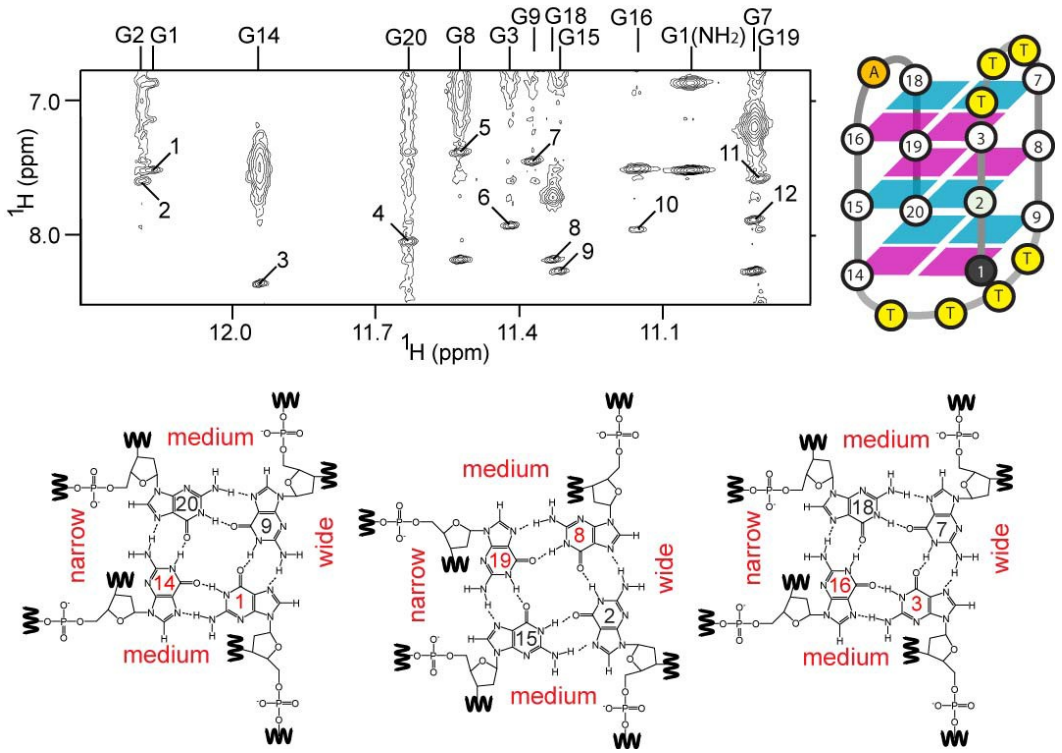


Figure S10. Exchangeable proton assignments for 5J05 at 5 °C. An expansion of a JR-NOESY spectrum (200 ms) in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 is shown, illustrating the dipolar connectivities between imino H1-H8 imino H1-H1. Peaks 1-12 are assigned as follows: (1) G1H1-G14H8, (2) G2H1-G8H8, (3) G14H1-G20H8, (4) G20H1-G9H8, (5) G8H1-G19H8, (6) G3H1-G16H8, (7) G9H1-G1H8, (8) G18H1-G7H8, (9) G15H1-G2H8, (10) G16H1-G18H8, (11) G19H1-G15H8, (12) G7H1-G3H8. These assignments allow for the formation of the hydrogen bond alignments shown in the chemical structures below.

Structural assignments for 5J4W in sodium

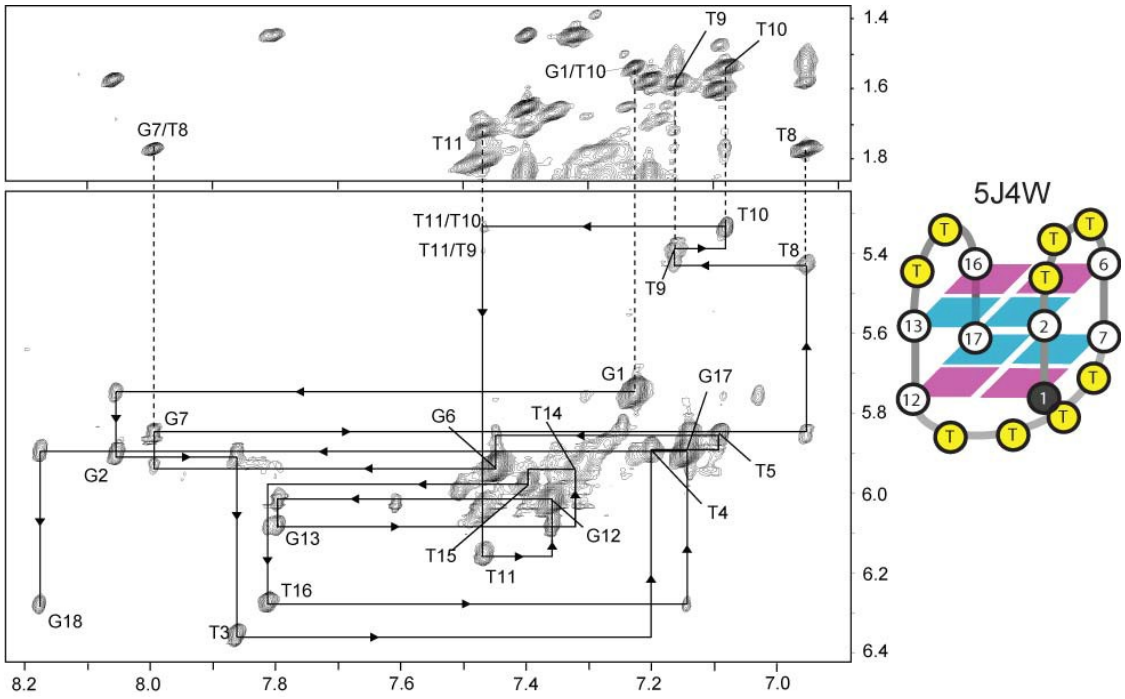


Figure S11. Non-exchangeable ^{1}H assignments for the 2(-l_wd+l_n) adopted by the DNA sequence 5J4W in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 and 20 ^{o}C. Expansions of ^{1}H-^{1}H NOESY spectra (20 ^{o}C) depicting anomeric-aromatic regions of the ^{1}H-^{1}H NOESY and showing labelled intraresidual H1′-H6/H8 and H6/H8-H2’/2’’ NOE interactions. Sequential correlations are denoted with lines. The inset contains assignments for the characteristic sequential connectivities (*Syn*G-*Anti*G-T-T-T-T) of the diagonal loop. Methyl-H8/H6 sections illustrate the characteristic connectivity between the aromatic H8 of the 5’-*Syn*G residue of the stem and the methyl of the third Thymine in the diagonal loop. Shown are also schematic representations of the topologies they adopt with 2’-deoxyguanosines of the stem in *syn* (pink) and *anti* (cyan) conformations, and yellow for thymines. A complete list of proton assignments is shown in **Table S5**.

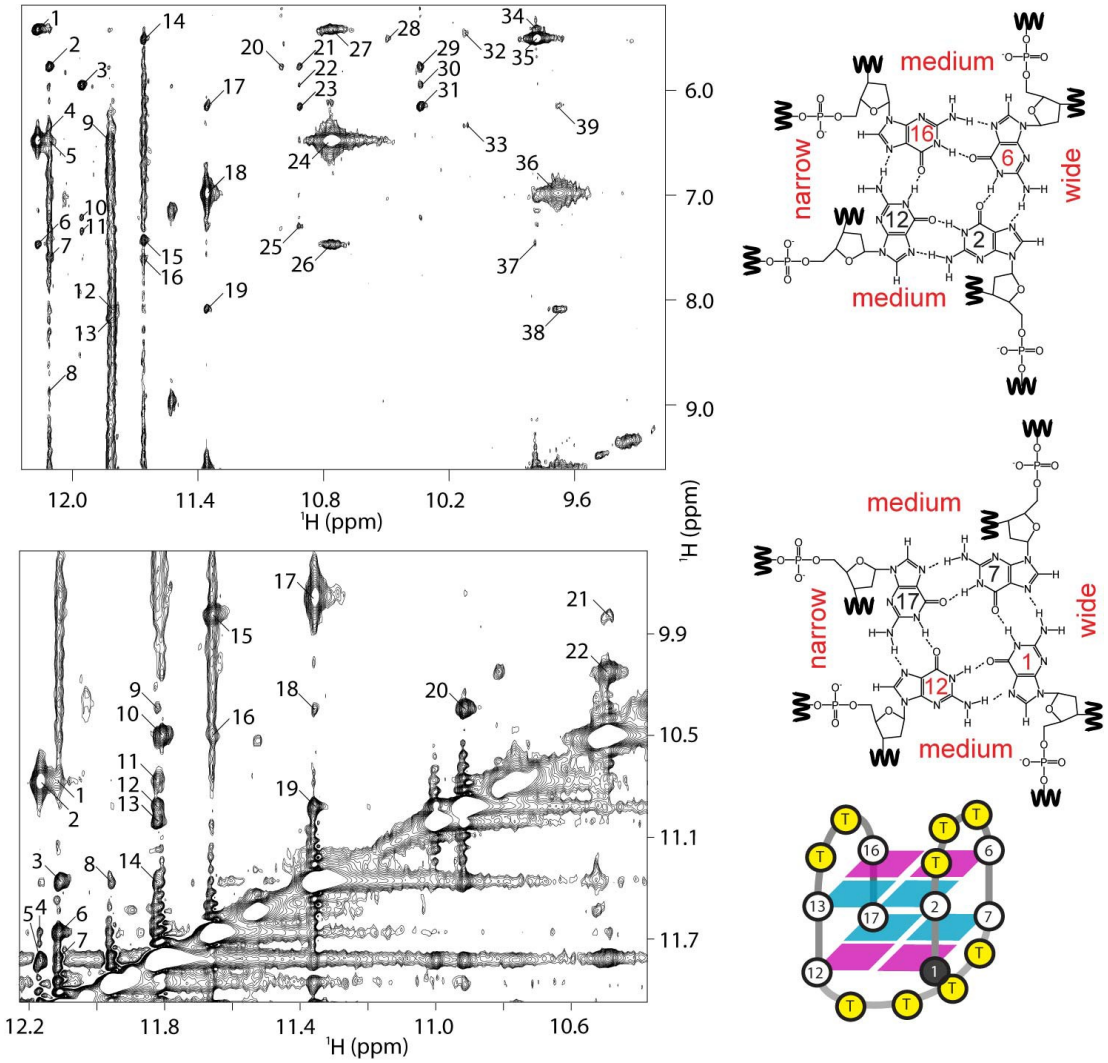


Figure S12. Expansions of a JR-NOESY (200 ms) spectrum of the DNA sequence 5J4W in 80 mM NaCl, 20 mM Na_2HPO_4/NaH_2PO_4, pH 6.8 in ^1H_2O at 5 °C. Top spectrum: illustration of NOE connectivities between imino and aromatic protons. Peaks 1-39 are assigned as follows: (1) G1H1-T10H1', (2) G2H1-T4H1', (3) G6H1-T5H1', (4) G1H1-G1H22, (5) G12H1-G1H22, (6) G1H1-G12H8, (7) G2H1-G6H8, (8) G12H1-G17H8, (9) G13H1-G1H22, (10) G6H1-T5H6, (11) G6H1-G16H8, (12) G17H1-G7H8, (13) G13H1-G2H8, (14) G7H1-T8H1', (15) G7H1-G1H8, (16) G7H1-T11H6, (17) G16H1-T15H1', (18) G16H1-G16H22, (19) G16H1-G13H8, (20) T4H3-T4H1', (21) T15H3-T4H1', (22) T15H3-T5H1', (23) T15H3-T15H1', (24) G1H21-G1H22, (25) T15H3-T15H6, (26) G1H21-G12H8, (27) G1H21-T10H1', (28) T8H3-T8H1', (29) T5H3-T4H1', (30) T5H3-T5H1', (31) T5H3-T15H1', (32) T9H3-T9H1', (33) T9H3-T11H1', (34) T10H3-T10H1' (35) T10H3-T8H1' (36) G16H21-G16H22 (37) T10H3-G1H8, (38) G16H21-G13H8, (39) G16H21-T15H1'. Bottom spectrum: illustration of imino-aromatic and imino-imino proton connectivities. Peaks 1-22 are assigned as follows: (1) G12H1-G1H21, (2) G1H1-G1H21, (3) G12H1-G16H1, (4) G1H1-G7H1, (5) G1H1-G13H1, (6) G2H1-G7H1, (7) G2H1-G13H1, (8) G6H1-G16H1, (9) G13H1-T5H3, (10) G17H1-T8H3, (11) G13H1-G1H21, (12) G13H1-T15H3, (13) G13H1-T4H3, (14) G13H1-G16H1, (15) G7H1-T10H3, (16) G7H1-T8H3, (17) G16H1-G16H21, (18) G16H1-T5H3, (19) G16H1-T15H3, (20) T15H3-T5H3, (21) T8H3-T10H3, (22) T8H3-T9H3. These assignments allow for the formation of the hydrogen bond alignments depicted in the chemical structures shown.

Structural assignments for 5J4P in sodium

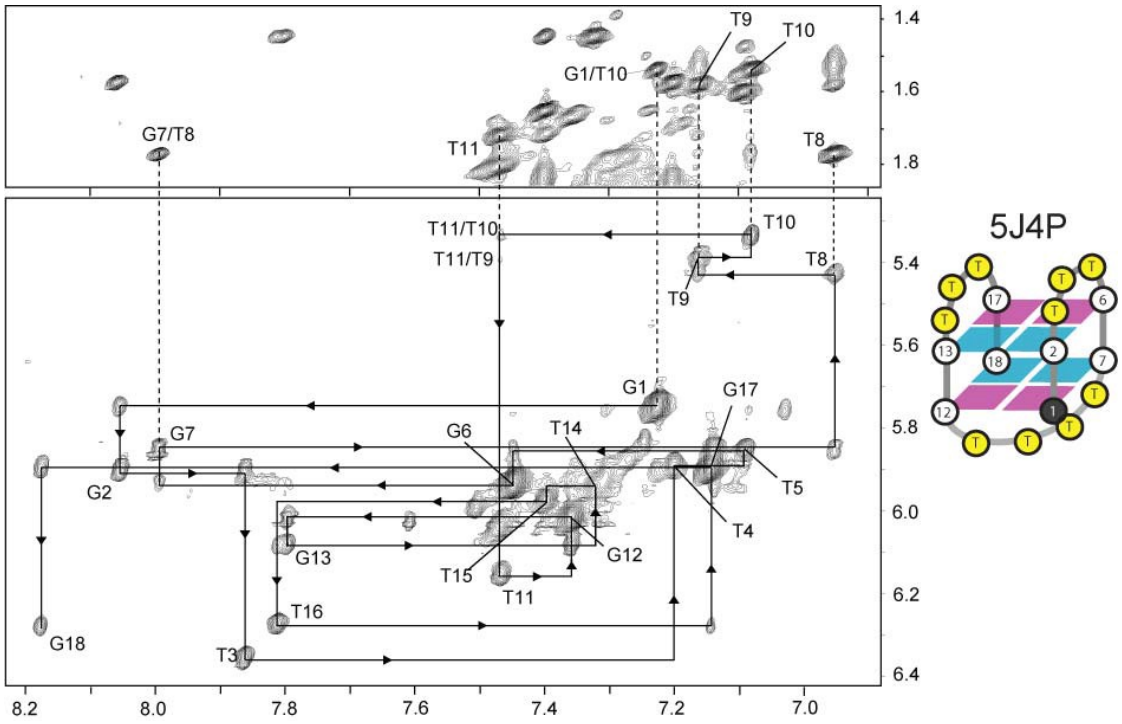


Figure S13. Non-exchangeable ^{1}H assignments for the 2(-l_wd+l_n) adopted by the DNA sequence 5J4P in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 and 20 ^{o}C. Expansions of ^{1}H-^{1}H NOESY spectra depicting anomeric-aromatic regions of the ^{1}H-^{1}H NOESY and showing labelled intraresidual H1′-H6/H8 and H6/H8-H2’/2’’ NOE interactions. Sequential correlations are denoted with lines. The inset contains assignments for the characteristic sequential connectivities (*Syn*G-*Anti*G-T-T-T-T) of the diagonal loop. Methyl-H8/H6 sections illustrate the characteristic connectivity between the aromatic H8 of the 5’-*Syn*G residue of the stem and the methyl of the third Thymine in the diagonal loop. Shown are also schematic representations of the topologies they adopt with 2’-deoxyguanosines of the stem in *syn* (pink) and *anti* (cyan) conformations, and yellow for thymines. A complete list of proton assignments is shown in **Table S6**.

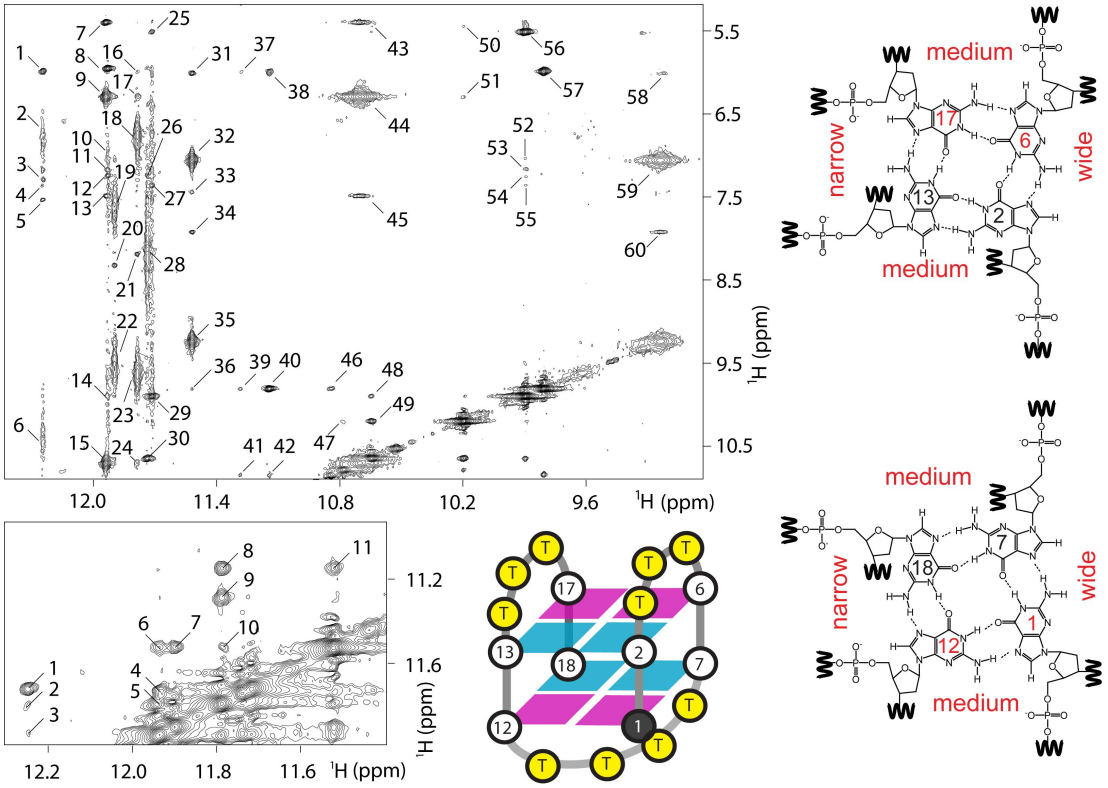


Figure S14. Expansions of a JR-NOESY (200 ms) spectrum of 5J4P collected in 80 mM NaCl, 20 mM Na_2HPO_4/NaH_2PO_4, pH 6.8 in ^1H_2O at 5 °C. Top spectrum illustrates NOE connectivities between imino and aromatic protons. Peaks 1-60 are assigned as follows: (1) G2H1-G2H1', (2) G2H1-G2H22, (3) G2H1-T4H6, (4) G2H1-G1H8, (5) G2H1-G6H8, (6) G2H1-G2H21, (7) G6H1-G10H1', (8) G6H1-T5H1', (9) G1H1-G1H22, (10) G6H1-G6H22, (11) G6H1-T5H6, (12) G6H1-G17H8, (13) G1H1-G12H8, (14) G6H1-G6H21, (15) G1H1-G1H21, (16) G13H1-G2H1', (17) G13H1-G1H22, (18) G13H1-G13H22, (19) G12H1-G12H22, (20) G12H1-G18H8, (21) G13H1-G2H8, (22) G12H1-G12H21, (23) G13H1-G13H21, (24) G13H1-G1H21, (25) G7H1-T8H1', (26) G18H1-G17H8, (27) G7H1-G1H8, (28) G18H1-G7H8, (29) G7H1-T10H3, (30) G18H1-T8H3, (31) G17H1-G17H1', (32) G17H1-G17H22, (33) G17H1-T14H6, (34) G17H1-G13H8, (35) G17H1-G17H21, (36) G17H1-T16H3, (37) T15H3-G2H1', (38) T14H3-G2H1', (39) T15H3-T16H3, (40) T14H3-T16H3, (41) T15H3-T5H3, (42) T14H3-T5H3, (43) G1H21-T10H1', (44) G1H21-G1H22, (45) G1H21-G12H8, (46) T15H3-T16H3, (47) T11H3-T9H3, (48) T8H3-T10H3, (49) T8H3-T9H3, (50) T9H3-T9H1', (51) T9H3-T11H1', (52) T10H3-T8H6, (53) T10H3-T10H6, (54) T10H3-T9H6, (55) T10H3-G1H8, (56) T10H1-T8H1', (57) T16H1-T4H1', (58) G17H21-G17H1', (59) G17H21-G17H22, (60) G17H21-G13H8. Bottom spectrum illustrates imino to imino proton connectivities. Peaks 1-12 are assigned as follows: (1) G2H1-G7H1, (2) G2H1-G13H1, (3) G2H1-G6H1, (4) G6H1-G18H1, (5) G1H1-G17H1, (6) G6H1-G17H1, (7) G12H1-G17H1, (8) G13H1-T14H3, (9) G13H3-T15H3, (10) G13H1-G17H1, (11) G17H1-T14H3. These assignments allow for the formation of the hydrogen bond alignments depicted in the chemical structures shown.

Structural assignments for 2M6V in sodium

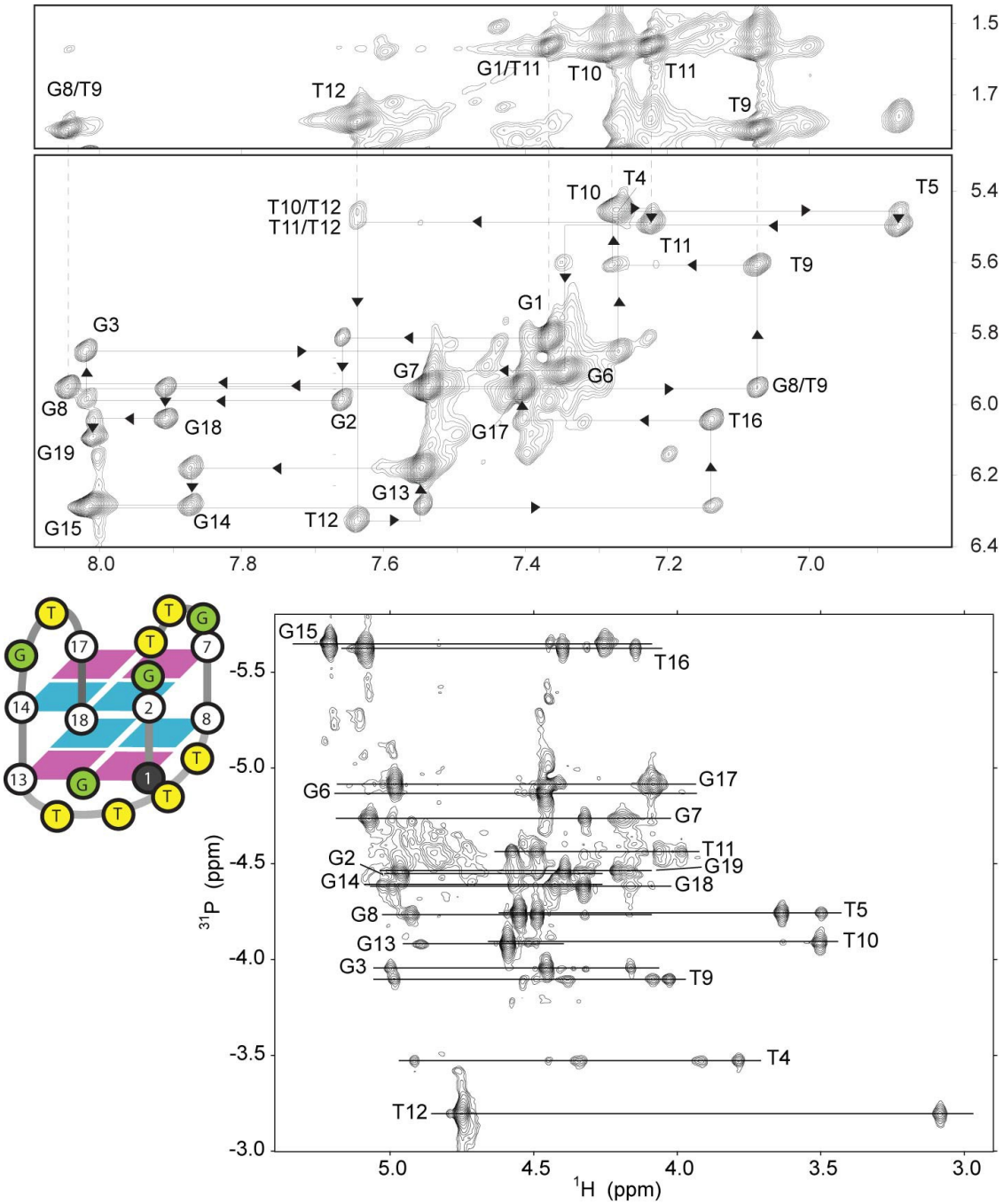


Figure S15. Non-exchangeable ^{1}H and ^{31}P assignments for the 2(-l_wd+l_n) adopted by the DNA sequence 2M6V in 20 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8, at 20 °C. The spectrum on top shows expansions of ^{1}H-^{1}H NOESY spectra (20 ^{o}C) depicting anomeric-aromatic regions of the ^{1}H-^{1}H NOESY and showing labelled intraresidual H1′-H6/H8 and H6/H8-H2'/2'' NOE interactions. Sequential correlations are denoted with lines. The inset contains assignments for the characteristic sequential connectivities (*Syn*G-*Anti*G-T-T-T-T) of the diagonal loop. Methyl-H8/H6 sections illustrate the characteristic connectivity between the aromatic H8 of the 5'-*Syn*G residue of the stem and the methyl of the third Thymine in the diagonal loop. Shown are also schematic representations of the topologies they adopt with 2'-deoxyguanosines of the stem in *syn* (pink) and *anti* (cyan) conformations, orange for adenines, and yellow for thymines. In the bottom spectrum sequential coupling correlations of the type H3'(i-1)-P(i)-H4'/H5'/H5'' in a [^{1}H-^{31}P] HSQC spectrum are shown. A complete list of proton assignments is shown in **Table S7**.

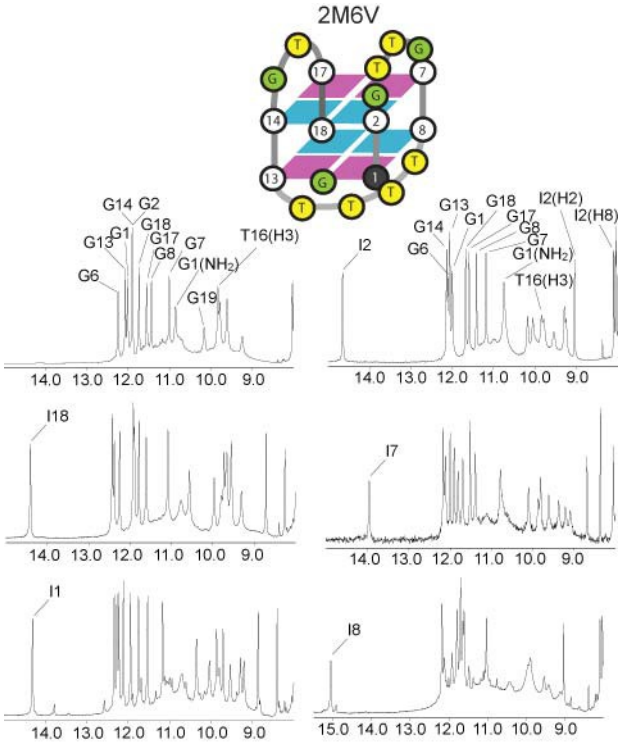


Figure S16. The exchangeable proton region expansions of 1D ^{1}H JR spectra of sequence 2M6V and of its inosine substitutions in 20 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8 at 5 ^{o}C. On the right, the [^{1}H-^{31}P] HSQC correlation spectrum of 2M6V showing observed individual (for each residue) strips, including correlations of the type H3'(i-1)-P(i)-H4'/H5'/H5'' that were the basis of confirming sequence specific assignments and deriving angular restraints.

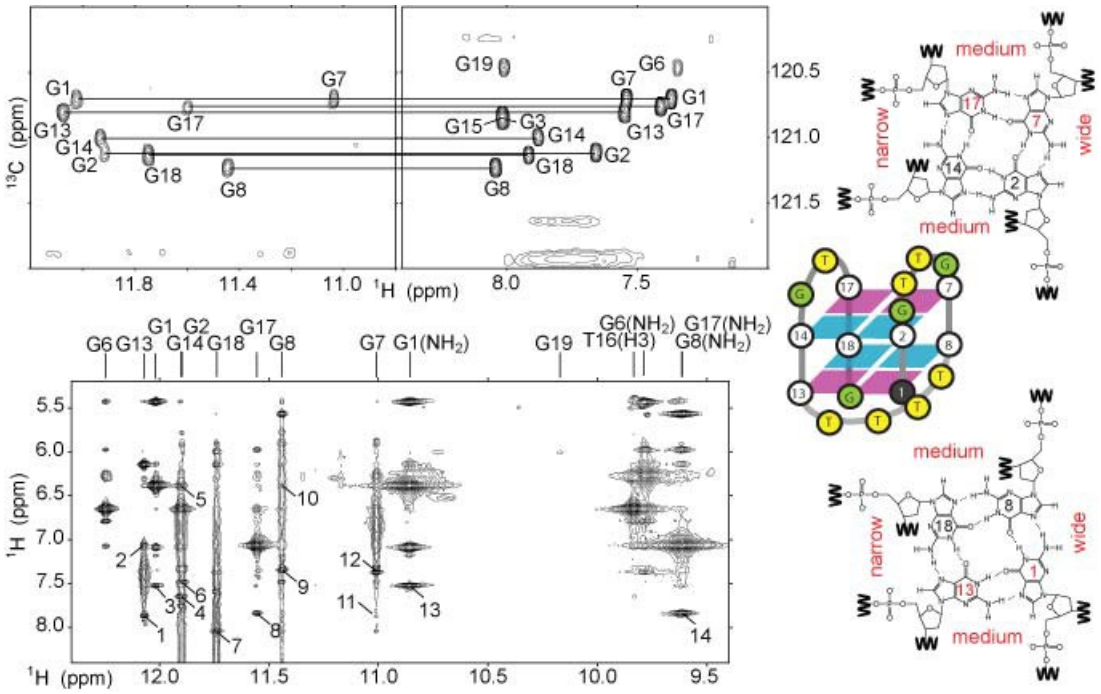


Figure S17. Solution NMR experiments providing evidence for the topology formed by the DNA sequence 2M6V in 20 mM NaCl, 4 mM NaH_2PO_4/Na_2HPO_4, pH 6.8. Top spectrum shows an

expansion of the natural abundance JR [^{1}H-^{13}C] HMBC spectrum illustrating the long range imino H1 to aromatic H8 coupling through the $^{13}C_5$ atom, in $^{1}H_2O$ at 10 °C. The bottom spectrum shows the exchangeable proton region of a JR-NOESY spectrum (250 ms) at 5° C, illustrating the dipolar connectivities between imino and amino exchangeable protons with aromatic protons (A). Peaks 1-14 are assigned as follows: (1) G13H1-G18H8, (2) G13H1-G17H22, (3) G1H1-G13H8, (4) G14H1-G2H8, (5) G14H1-G1H22, (6) G2H1-G7H8, (7) G18H1-G8H8, (8) G17H1-G14H8, (9) G8H1-G1H8 (10), G8H1-G1H22, (11) G7H1-G18H8, (12) G7H1-G17H8, (13) G1H21-G13H8 and (14) G17H21-G14H8. These assignments allow for the formation of the hydrogen bond alignments depicted in the chemical structures shown.

C. IDENTIFICATION OF TOPOLOGY

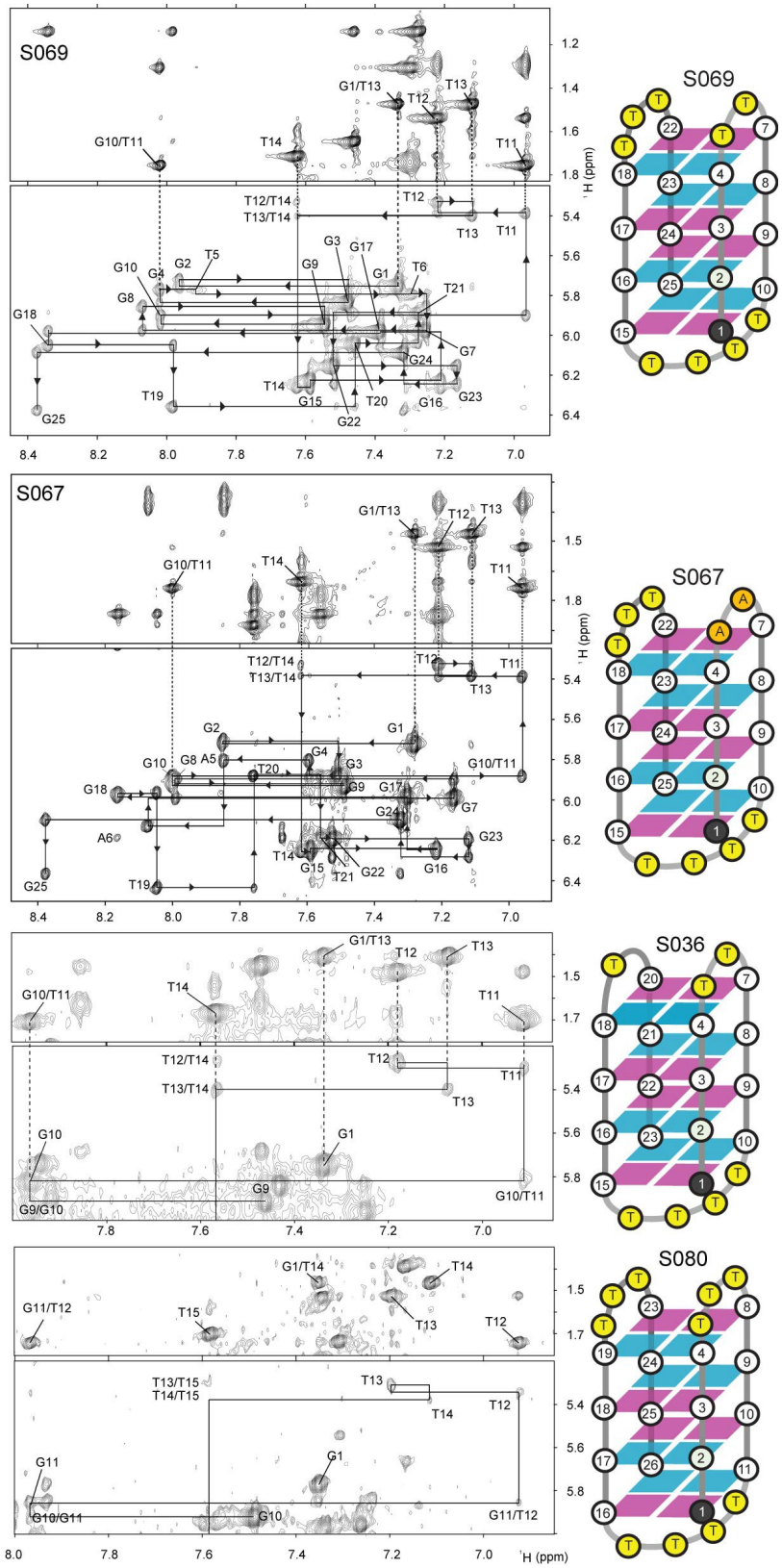


Figure S18. Anomeric-aromatic regions of the ^{1}H-^{1}H NOESY (tm 250 ms, 20 ^{o}C) spectra of DNA sequences S069, S067, S036 and S080 in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8. Intraresidual H1′-H6/H8 and H6/H8-H2’/2’’ NOE interactions are labelled and sequential correlations are denoted with lines. The inset contains assignments for the characteristic sequential connectivities

(*Syn*G-*Anti*G-T-T-T-T) of the diagonal loop. Methyl-H8/H6 sections illustrate the characteristic connectivity between the aromatic H8 of the 5'-*Syn*G residue of the stem and the methyl of the third Thymine in the diagonal loop. Shown are also schematic representations of the topologies they adopt with 2'-deoxyguanosines of the stem in *syn* (magenta) and *anti* (cyan) conformations.

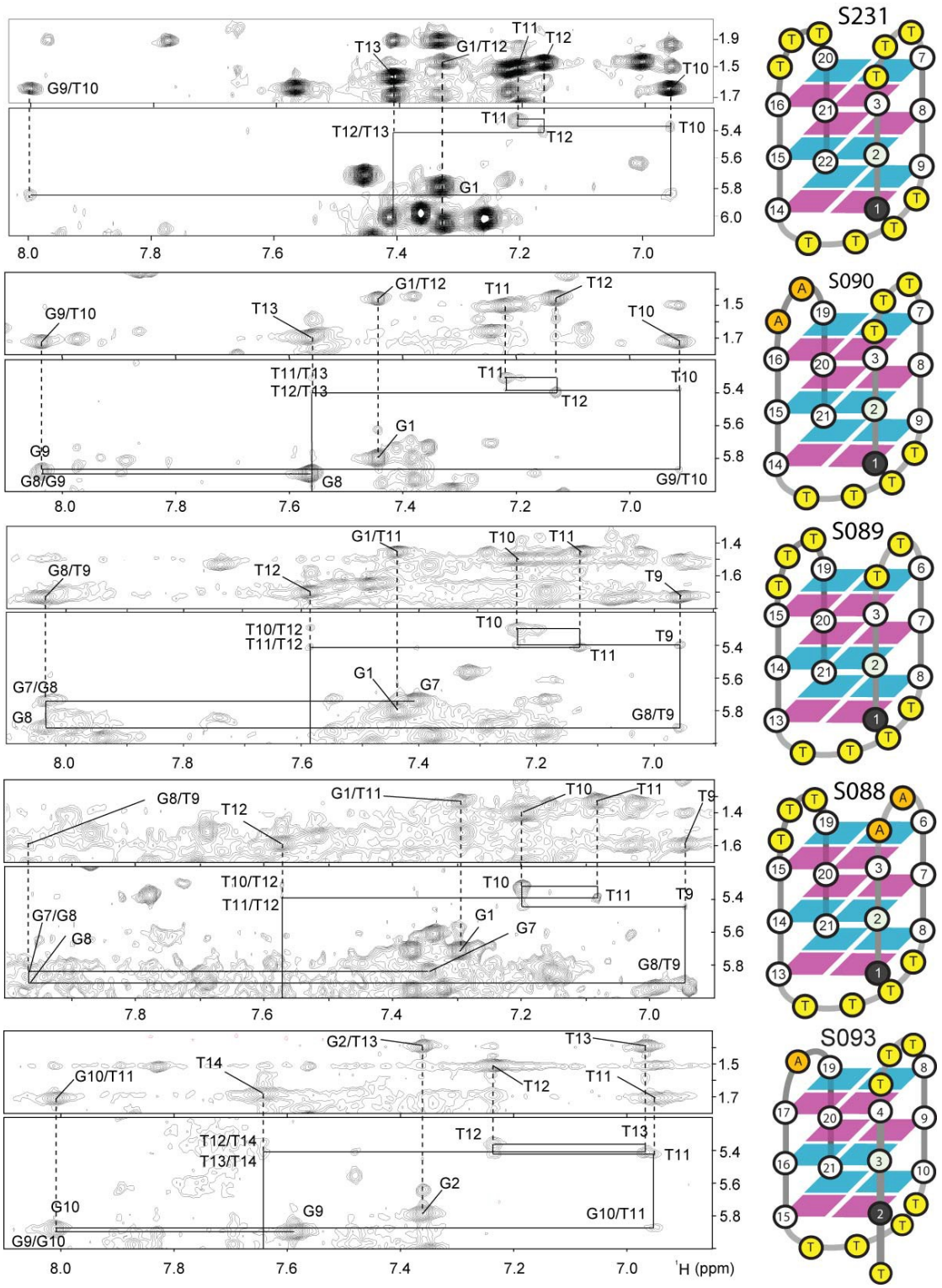


Figure S19. Expansions of the anomeric-aromatic regions of the ^{1}H-^{1}H NOESY (mixing time 250 ms, 20 ^{o}C) spectra of DNA sequences S231, S090, S089, S088, and S093 in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8. Intraresidual H1′-H6/H8 NOE interactions are labelled and sequential

correlations are denoted with lines. The inset contains assignments for the characteristic sequential connectivities (*Syn*G-*Anti*G-T-T-T-T) of the diagonal loop. Methyl-H8/H6 sections illustrate the characteristic connectivity between the aromatic H8 of the 5'-*Syn*G residue of the stem and the methyl of the third Thymine in the diagonal loop. Shown are also schematic representations of the topologies they adopt with 2'-deoxyguanosines of the stem in *syn* (magenta) and *anti* (cyan) conformations.

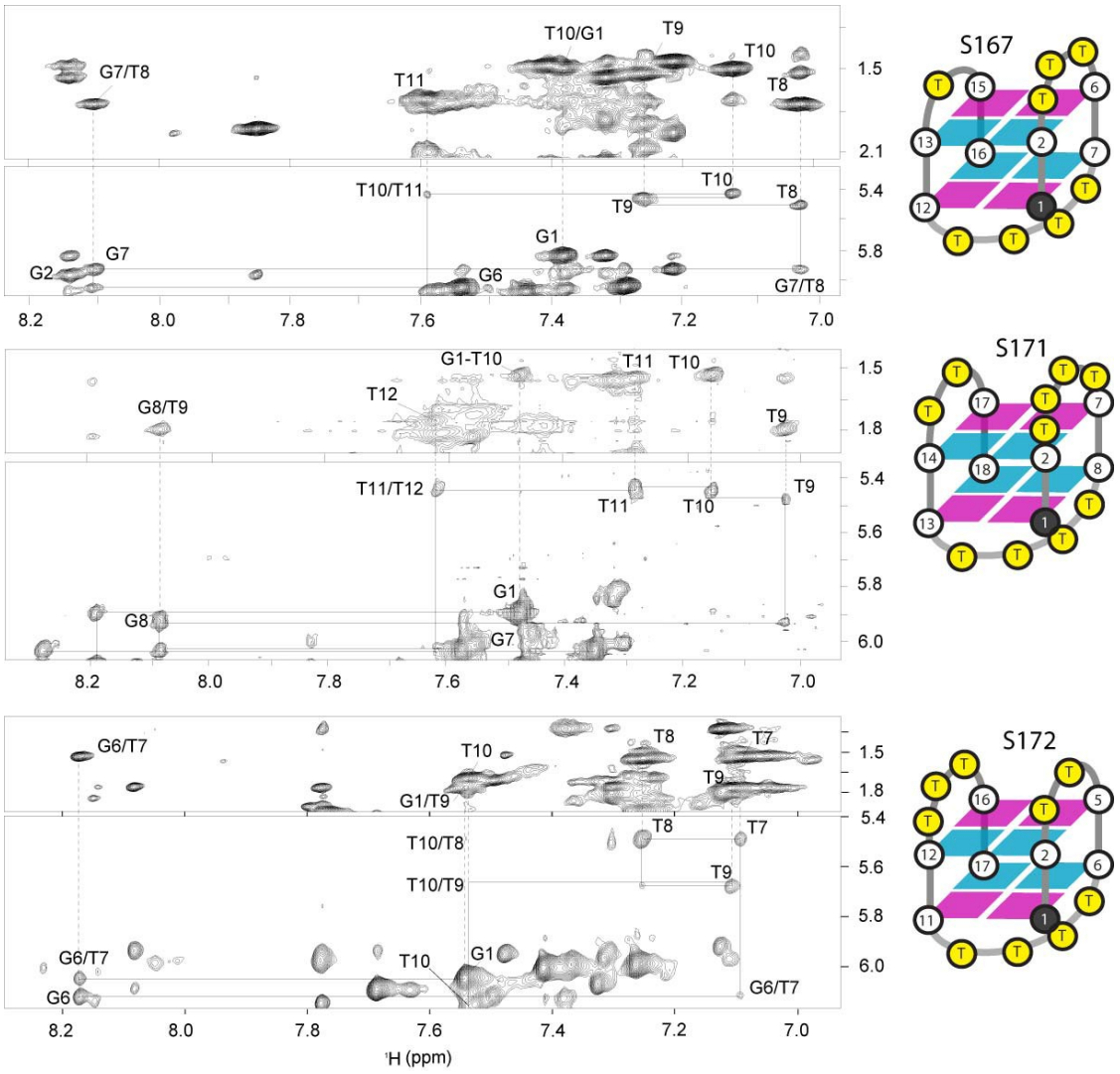


Figure S20. Anomeric-aromatic regions of the ^{1}H-^{1}H NOESY (tm 250 ms, 20 ^{o}C) spectra of DNA sequences S167, S171, and S172 in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8. Intraresidual H1′-H6/H8 NOE interactions are labelled and sequential correlations are denoted with lines. The inset contains assignments for the characteristic sequential connectivities (*Syn*G-*Anti*G-T-T-T-T) of the diagonal loop. Methyl-H8/H6 sections illustrate the characteristic connectivity between the aromatic H8 of the 5'-*Syn*G residue of the stem and the methyl of the third thymine in the diagonal loop. Shown are also schematic representations of the topologies they adopt with 2'-deoxyguanosines of the stem in *syn* (magenta) and *anti* (cyan) conformations.

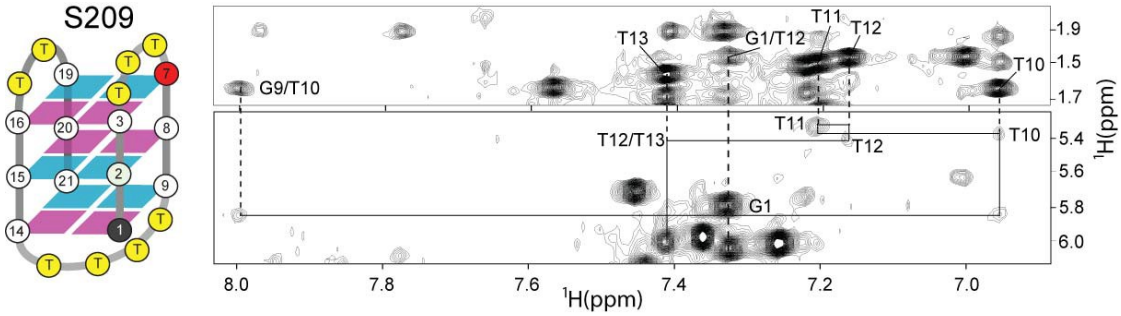


Figure S21. Use of riboguanosines to induce folding of the 3(-l_wd+l_n) topology. Schematics of the topology adopted by S209 in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8. Guanosines in tetrads comprising the quadruplex stem are displayed as pink (*syn*G) and blue (*anti*G) squares, with numbered white (deoxyriboguanosine) or red (riboguanosine) circles indicating their position in the primary sequence. Thymines of each loop are shown as yellow circles. Expansions of proton NOESY spectrum at 250 ms mixing time and 20 °C, illustrating the presence of the characteristic sequential connectivities (*syn*G-*anti*G-T-T-T-T) of the diagonal loop in the 3(-l_wd+l_n) topology. For each sequential assignment the characteristic connectivity between the aromatic H8 of the 5'-*Syn*G residue of the stem and the methyl of the third Thymine in the diagonal loop is illustrated.

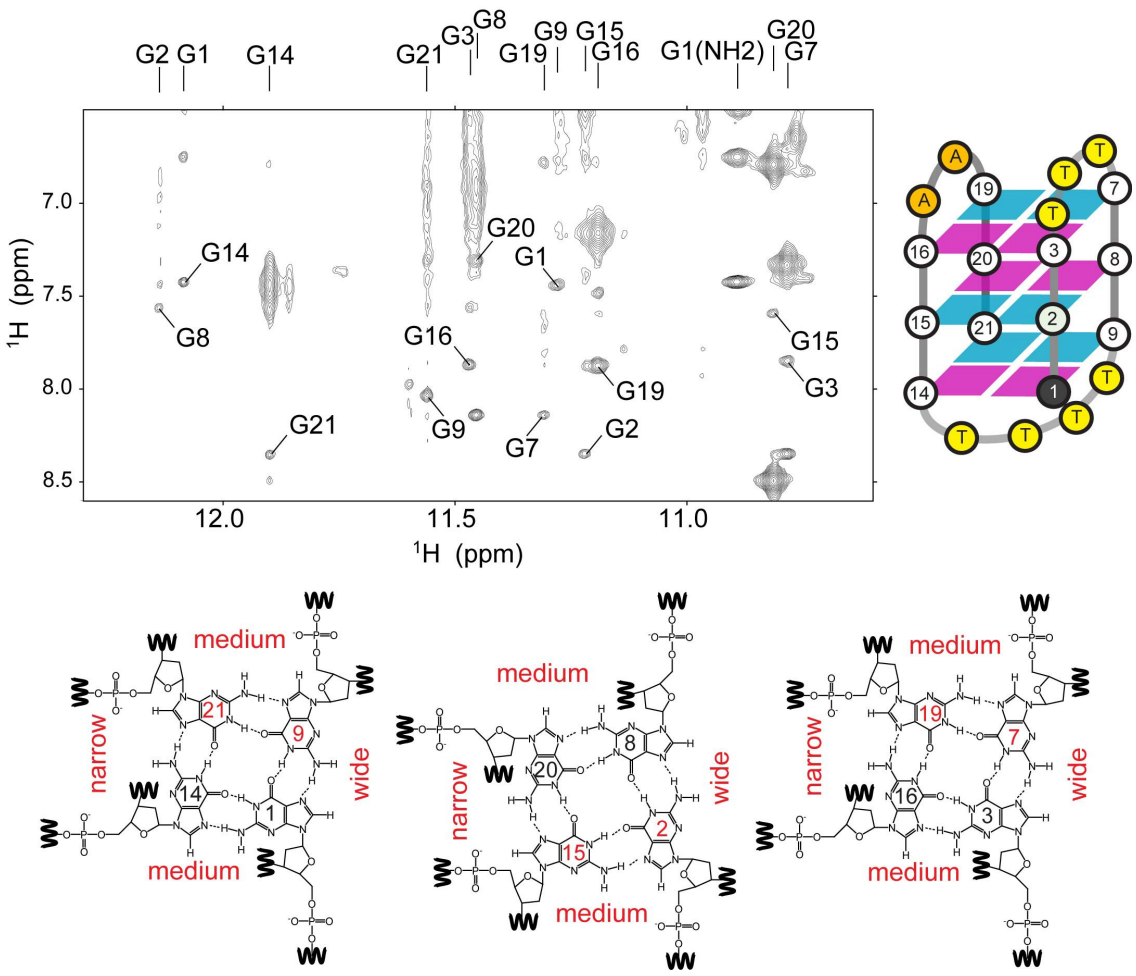


Figure S22. Exchangeable proton assignments for S090 in 80 mM NaCl, 20 mM NaH_2PO_4/Na_2HPO_4, pH 6.8. An expanded [^{1}H-^{1}H] JR-NOESY spectrum (200 ms) of S090 at 5° C, depicts NOE connectivities between imino (H1) exchangeable protons and aromatic (H8) protons. Sequential connectivities and those with loop residues are not indicated. These assignments allow for the hydrogen bond alignments depicted in the chemical structures below defining the topology shown.

D. NMR CHEMICAL SHIFTS TABLES

Table S1. Proton chemical shifts for sequence 2MFT. Exchangeable-proton chemical shifts were obtained at 5° C while non-exchangeable proton chemical shifts were obtained at 20° C. H5'/H5'' protons are not stereo-specifically assigned.

Residue	H1/H3	H21	H22	H8/H6	CH_3	H1'	H2'	H2''	H3'	H4'	H5'	H5''
G1	11.736	10.184	6.215	7.253	-	5.766	2.894		5.064	4.308	4.972	4.014
G2	11.686	11.514	6.109	8.029	-	5.295	2.514	2.294	5.031	4.146	-	-
G3	11.514	-	-	7.976	-	6.141	2.820	2.581	4.940	4.367	4.063	
T4	-	-	-	7.073	1.753	5.614	2.274	1.509	4.503	4.007	-	-
T5	-	-	-	7.227	1.557	5.431	1.973	1.842	4.500	3.462	3.601	
T6	9.841	-	-	7.082	1.473	5.300	2.290	1.703	4.668	4.359	3.984	3.883
T7	-	-	-	7.566	1.660	6.236	2.367	2.073	4.524	3.183	3.014	2.790
G8	12.197	8.842	6.857	7.292	-	6.185	3.382	3.020	4.949	4.448	4.144	4.012
G9	11.119	9.395	6.776	7.395	-	5.982	3.227	2.864	5.020	4.443	4.317	
G10	11.574	-	-	8.199		6.284	2.663	2.447	4.766	4.298	4.213	
T11	-	-	-	7.900	2.003	6.526	2.793	2.422	5.072	4.382	-	-
G12	11.590	10.822	5.968	7.175	-	5.886	3.298		4.971	4.306	3.927	3.834
G13	11.669	-	-	7.346	-	5.667	2.508		5.034	4.140	-	-
G14	11.619	-	-	7.944	-	5.911	2.685	2.511	4.929	4.439	-	-
T15	-	-	-	6.931	1.685	5.494	2.143	1.365	4.465	3.898	4.322	4.036
T16	-	-	-	7.197	1.536	5.397	2.001	1.859	4.508	3.364	3.541	3.478
T17	9.896	-	-	7.055	1.425	5.481	2.297	1.672	4.655	4.479	4.035	3.934
T18	-	-	-	7.610	1.686	6.214	2.362	2.080	4.538	3.175	2.870	2.773
G19	11.571	9.549	7.299	7.528	-	6.220	3.579	3.033	4.930	4.728	4.138	
G20	11.565	9.434	6.722	7.888	-	6.211	2.923	2.565	5.087	4.585	4.385	4.288
G21	11.086	9.082	6.615	7.935	-	6.628	2.827	2.736	5.174	4.517	4.378	

Table S2. Proton and phosphorous (20 °C) chemical shifts for sequence 2M6W in 20 mM NaPi (pH 6.8) aqueous buffer. Exchangeable-proton chemical shifts were obtained at 5° C while non-exchangeable proton chemical shifts were obtained at 20° C. H5'/H5'' protons are not stereo-specifically assigned.

Residue	H1/H3	H21	H22	P	H8/H6	H2	CH_3	H1'	H2'	H2''	H3'	H4'	H5'	H5''
G1	11.73	10.9	6.087	-	7.308	-	-	5.738	2.388	2.785	4.858	4.281	3.91	3.789
G2	11.88	9.691	-	-4.016	7.908	-	-	5.698	2.531	2.812	5.071	4.227	4.127	4.148
G3	10.58	9.685	6.766	-3.425	7.572	-	-	5.863	2.745	2.482	4.903	4.257	4.053	
G4	10.9	9.314	7.046	-4.171	8.008	-	-	5.768	2.568		4.882	4.49	4.062	
T5	-	-	-	-3.762	7.770	-	2.008	5.767	2.31	1.866	4.699	4.315	4.27	4.042
T6	-	-	-	-5.209	7.084	-	1.215	5.699	1.722	1.856	4.298	2.369	3.595	3.301
G7	11.69	10.06	6.305	-3.824	7.272	-	-	5.99	3.532	2.707	4.758	4.589	4.283	3.865
G8	11.41	-	-	-4.27	8.054	-	-	5.857	2.751	3.053	5.084	4.318	4.202	
G9	11.13	9.699	6.834	-3.889	7.546	-	-	5.951	2.891	2.625	5.04	4.354	4.192	
G10	10.95	9.955	-	-3.996	8.021	-	-	5.903	2.588	2.713	4.968	4.485	4.038	
T11	-	-	-	-4.529	6.971	-	1.751	5.377	1.300	2.071	4.45	3.851	3.79	3.931
T12	-	-	-	-4.337	7.219	-	1.538	5.325	2.089	1.886	4.539	3.463	3.554	3.391
T13	9.512	-	-	-3.866	7.113	-	1.465	5.407	1.602	2.202	4.669	4.447	4.069	3.932
T14	-	-	-	-5.343	7.621	-	1.706	6.252	2.066	2.381	4.558	3.162	2.639	2.879
G15	11.8	9.696	7.509	-3.984	7.597	-	-	6.22	3.557	2.878	4.943	4.717	4.496	4.142
G16	10.99	9.263	6.378	-	7.215	-	-	6.232	2.572	2.753	5.057	4.516	4.328	
G17	11.36	8.626	6.364	-4.057	7.33	-	-	5.933	3.176	2.433	5.08	4.485	4.056	4.325
G18	11.4	9.684	-	-3.73	7.717	-	-	6.122	2.959	2.714	5.124	4.457	4.109	4.252
A19	-	-	-	-2.153	8.486	7.766	-	6.419	2.951	2.801	5.196	4.529	4.367	4.336
A20	-	-	-	-3.229	8.054	7.912	-	6.263	2.971	2.808	5.115	4.488	4.078	4.241
G21	10.54	9.685	6.182	-2.887	7.539	-	-	6.211	3.514	3.018	5.031	4.443	4.493	4.298
G22	11.1	9.38	6.397	-3.579	7.179	-	-	6.214	2.520	2.616	4.998	4.519	4.366	
G23	11.11	8.577	6.668	-3.915	7.32	-	-	6.095	3.600	2.993	5.119	4.261	4.573	4.73
G24	11.44	9.697	6.835	-3.777	8.361	-	-	6.376	2.768	2.527	4.839	4.349	4.106	4.263

Table S3. Proton and phosphorous (20 °C) chemical shifts for sequence 5J6U in 20 mM NaPi (pH 6.8) aqueous buffer. Exchangeable-proton chemical shifts were obtained at 5° C at 500 MHz, while non-exchangeable proton chemical shifts were obtained at 20° C at 900 MHz. H5'/H5'' protons are not stereo-specifically assigned.

Residue	H1/H3	H21	H22	P	H8/H6	H2	CH_3	H1'	H2'	H2''	H3'	H4'	H5'	H5''
G1	11.52	10.19	5.82	-	7.30	-	-	6.03	3.08	2.83	4.95	4.43	3.88	3.89
G2	11.51	9.25	6.03	-2.23	7.97	-	-	6.19	2.84	2.70	5.09	4.47	4.12	4.41
G3	10.87	8.92	-	-3.55	7.24	-	-	5.96	2.82	3.21	5.08	4.35	4.39	4.11
G4	11.35	9.60	6.16	-5.81	8.13	-	-	6.10	2.52	2.52	2.80	5.12	4.47	4.22
T5	-	-	-	-4.25	7.34	-	1.91	6.28	2.39	2.51	4.73	4.25	4.15	4.12
T6	-	-	-	-2.66	6.53	-	0.99	5.36	2.21	1.45	4.50	3.91	3.54	3.84
T7	-	-	-	-3.41	6.90	-	0.83	6.03	2.17	1.82	4.55	4.17	3.57	2.76
G8	11.66	9.46	5.86	-	7.46	-	-	6.00	3.14	2.96	4.72	4.39	4.32	4.17
G9	11.45	9.69	6.08	-2.06	7.99	-	-	6.17	2.79	2.63	5.04	4.48	4.31	4.09
G10	10.99	8.74	6.03	-3.62	7.24	-	-	5.9	2.65	3.04	5.03	4.49	4.10	4.32
G11	11.34	-	-	-2.86	7.91	-	-	5.97	2.68	2.53	4.99	4.44	4.18	4.13
T12	-	-	-	-3.76	7.10	-	1.78	5.6	2.12	1.82	4.62	4.33	3.90	4.06
T13	-	-	-	-2.48	7.32	-	1.67	5.67	2.03	1.88	4.50	3.70	-	-
T14	-	-	-	-3.66	7.18	-	1.41	5.73	2.04	1.82	-	4.35	4.34	3.82
T15	-	-	-	-	7.21	-	1.67	6.06	2.43	1.87	-	-	-	-
G16	11.28	9.32	7.22	-	7.59	-	-	6.14	3.04	3.50	4.85	4.64	4.12	4.11
G17	11.39	9.39	6.02	-1.92	7.87	-	-	6.16	2.71	2.57	5.04	4.46	4.23	4.08
G18	11.09	8.43	5.97	-3.70	7.23	-	-	6.03	3.30	2.88	5.16	4.57	4.36	4.07
G19	11.00	-	-	-5.96	8.13	-	-	6.2	2.62	2.80	5.15	4.94	4.40	4.42
A20	-	-	-	-4.64	8.24	8.10	-	6.32	2.54	2.70	4.94	4.44	4.18	4.16
A21	-	-	-	-5.43	7.68	7.64	-	6.12	2.51	3.02	4.89	4.34	3.97	3.87
G22	10.77	9.34	6.63	-	7.08	-	-	5.73	2.84	3.02	4.89	4.36	4.33	4.23
G23	-	-	-	-2.06	7.68	-	-	6.09	2.66	2.41	4.99	4.48	4.05	4.06
G24	11.02	8.61	6.17	-3.77	7.22	-	-	6.00	2.83	3.25	5.07	4.59	4.45	4.05
G25	11.29	8.73	6.04	-	8.19	-	-	6.22	2.64	2.41	4.75	4.26	4.19	4.26

Table S4. Proton (20 °C) chemical shifts for sequence 5J05 in 20 mM NaPi (pH 6.8) aqueous buffer. Exchangeable-proton chemical shifts were obtained at 5° C while non-exchangeable proton chemical shifts were obtained at 20° C. H5'/H5'' protons are not stereo-specifically assigned

Residue	H1/H3	H21	H22	H8/H6	H2	CH_3	H1'	H2'	H2''	H3'	H4'	H5'	H5''
G1	12.161	11.039	6.408	7.358	-	-	5.761	2.824		4.871	4.156	-	-
G2	12.182	10.125	6.997	8.15	-	-	5.566	2.568	2.505	5.151	4.195	4.117	4.198
G3	11.414	10.363	6.594	7.827	-	-	6.311	2.789		5.708	4.451	4.289	4.17
T4	-	-	-	6.673	-	1.46	5.147	1.408	2.1	4.383	3.775	3.627	
T5	-	-	-	7.495	-	1.814	6.105	2.328		4.657	4.106	4.029	3.901
T6	-	-	-	6.675	-	0.694	5.802	1.658	1.949	4.486	3.869	3.693	
G7	10.905	9.879	7.186	8.091	-	-	5.78	3.267	2.596	4.575	4.142	-	-
G8	11.521	9.928	6.931	7.535	-	-	5.851	2.617		4.959	4.189	4.102	
G9	11.368	10.205	6.205	7.949	-	-	5.855	2.477	2.625	4.886	4.423	4.421	3.988
T10	-	-	-	6.881	-	1.69	5.362	1.177	2.005	4.401	3.814	3.341	3.344
T11	-	-	-	7.183	-	1.486	5.306	2.084	1.852	4.497	3.498	3.447	3.495
T12	-	-	-	7.076	-	1.424	5.404	1.565	2.218	4.649	4.398	2.657	4.264
T13	-	-	-	7.569	-	1.673	6.21	2.019	2.312	4.473	3.044	2.731	3.044
G14	11.939	9.491	7.514	7.438	-	-	6.09	3.419	2.895	4.849	4.254	4.06	4.04
G15	11.313	9.813	6.614	7.511	-	-	6.074	2.599	2.792	5.097	4.425	-	-
G16	11.152	9.976	6.414	7.866	-	-	6.246	2.954	2.59	5.026	4.244	3.777	3.629
A17	-	-	-	7.446	7.664	-	5.927	1.959	2.76	4.882	4.293	4.161	
G18	11.326	10.416	5.848	7.887	-	-	5.992	2.516	2.784	4.891	4.386	4.015	
G19	10.893	8.673	6.641	7.32	-	-	5.995	3.544	2.825	5.035	4.267	4.159	4.251
G20	11.628	9.292	6.999	8.263	-	-	6.283	2.703	2.418	4.252	4.165	-	-

Table S5. Proton (20 °C) chemical shifts for sequence 5J4W in 100 mM Na+ (pH 6.8) aqueous buffer. Exchangeable-proton chemical shifts were obtained at 5° C while non-exchangeable proton chemical shifts were obtained at 20° C. H5'/H5'' protons are not stereo-specifically assigned.

Residue	H1/H3	H21	H22	H8/H6	CH_3	H1'	H2'	H2''	H3'	H4'	H5'	H5''
G1	12.17	10.76	6.48	7.39	-	5.86	2.74	2.81	4.99	4.19	4.18	4.19
G2	12.11	-	-	8.14	-	6.01	2.51	2.80	5.09	4.23	-	-
T3	-	-	-	7.84	1.97	6.36	2.40	2.61	4.38	4.10	3.08	3.09
T4	11.00	-	-	7.34	1.65	5.80	1.89	2.24	4.62	4.06	3.52	3.10
T5	10.33	-	-	7.22	1.45	5.95	2.38	2.23	4.59	3.57	4.10	
G6	11.96	-	-	7.60	-	6.08	3.42	2.78	4.37	4.19	-	-
G7	11.66	-	-	8.06	-	5.94	2.65	2.76	5.00	4.38	4.09	4.08
T8	10.49	-	-	7.04	1.78	5.50	1.45	2.25	4.45	4.01	3.61	3.60
T9	10.12	-	-	7.28	1.57	5.48	2.08	1.94	4.56	3.62	3.54	3.61
T10	9.78	-	-	7.16	1.54	5.43	1.75	2.37	4.46	-	-	-
T11	-	-	-	7.61	1.74	6.30	2.13	2.40	4.53	3.15	2.80	2.94
G12	12.11	-	-	7.45	-	6.10	3.45	2.98	4.99	-	4.14	4.12
G13	11.82	-	-	8.05	-	6.19	2.89	2.48	5.12	4.40	4.33	4.32
T14	-	-	-	7.82	2.00	6.24	2.58	2.24	-	4.32	-	-
T15	10.91	-	-	7.30	1.26	6.15	2.65	2.09	4.93	-	-	-
G16	11. 36	9.67	6.98	7.34	-	6.05	3.53	2.96	4.97	4.28	-	-
G17	11.81	-	-	8.27	-	6.31	2.73	2.48	-	4.29	-	-

Table S6. Proton (20 °C) chemical shifts for sequence 5J4P in 100 mM Na+ (pH 6.8) aqueous buffer. Exchangeable-proton chemical shifts were obtained at 5° C while non-exchangeable proton chemical shifts were obtained at 20° C. H5'/H5'' protons are not stereo-specifically assigned.

Residue	H1/H3	H21	H22	H8/H6	CH_3	H1'	H2'	H2''	H3'	H4'	H5'	H5''
G1	11.94	10.71	6.29	7.33	-	5.85	2.88	2.79	4.99	4.36	-	-
G2	12.25	10.39	6.83	8.16	-	6.01	2.69	2.68	5.09	4.34	4.21	4.19
T3	-	-	-	7.96	2.00	6.45	2.63	2.51	4.42	4.10	4.10	4.11
T4	-	-	-	7.30	1.57	5.99	2.12	1.97	4.55	4.10	3.92	3.92
T5	10.85	-	-	7.19	1.60	5.95	2.38	2.38	4.78	3.79	3.57	2.95
G6	11.93	9.90	7.03	7.55	-	6.03	3.33	3.32	4.24	4.17	4.10	3.79
G7	11.72	-	-	8.10	-	5.95	2.76	2.66	4.99	4.47	-	-
T8	10.65	-	-	7.06	1.77	5.53	2.27	1.53	4.45	4.03	3.51	3.62
T9	10.20	-	-	7.26	1.58	5.49	1.93	2.02	4.54	4.54	3.63	3.52
T10	9.90	-	-	7.18	1.54	5.43	2.34	2.06	4.71	4.43	3.99	3.98
T11	10.79	-	-	7.57	1.72	6.25	2.40	2.11	4.55	-	3.19	2.91
G12	11.82	9.53	7.59	7.09	-	6.12	3.18	3.42	4.96	4.72	4.13	4.14
G13	9.64	6.76	11.79	7.68	-	6.18	2.55	2.62	5.18	4.44	-	-
T14	11.15	-	-	7.43	1.44	6.04	2.50	2.40	4.93	4.19	-	-
T15	11.29	-	-	7.50	1.65	6.07	2.52	2.18	4.82	4.28	3.97	3.96
T16	9.81	-	-	7.91	2.08	6.37	2.50	2.23	4.88	4.51	4.27	4.14
G17	11.52	9.24	7.06	7.25	-	5.99	3.58	3.59	4.97	4.44	4.32	4.31
G18	11.74	-	8.34	8.28	-	6.38	2.52	2.78	-	-	-	-

Table S7. Proton and phosphorous (20 °C) chemical shifts for sequence 2M6V in 20 mM NaPi (pH 6.8) aqueous buffer. Exchangeable-proton chemical shifts were obtained at 5° C while non-exchangeable proton chemical shifts were obtained at 20° C. H5'/H5'' protons are not stereo-specifically assigned.

Residue	H1/H3	H21	H22	P	H8/H6	CH_3	H1'	H2'	H2''	H3'	H4'	H5'	H5''
G1	12.02	10.86	6.381	-	7.369	-	5.811	2.728	2.56	4.963	4.357	4.027	3.877
G2	11.9	-	-	-4.450	7.663	-	5.989	2.566	2.248	5.002	4.401	4.141	
G3	-	-	-	-3.952	8.023	-	5.85	2.646	2.521	4.922	4.462	4.332	4.162
T4	-	-	-	-3.470	7.271	1.867	5.455	2.023	1.92	4.558	3.789	4.35	3.928
T5	6.649	-	-	-4.243	6.876	1.274	5.498	2.332	1.757	4.467	3.64	3.499	2.635
G6	12.25	9.785	6.271	-4.871	7.347	-	5.906	2.828	2.627	5.079	4.194	4.355	3.987
G7	11.01	9.791	6.811	-4.732	7.54	-	5.95	3.161	2.654	4.934	4.33	4.232	4.184
G8	11.44	9.614	-	-4.232	8.048	-	5.951	2.718	2.639	4.991	4.493	4.397	4.089
T9	-	-	-	-3.895	7.076	1.795	5.606	2.232	1.525	4.513	4.034	3.631	3.527
T10	-	-	-	-4.089	7.283	1.576	5.45	2.078	1.913	4.579	3.51	3.639	3.55
T11	7.09	-	-	-4.562	7.227	1.56	5.489	2.419	1.758	4.753	4.493	4.087	3.998
T12	-	-	-	-3.192	7.64	1.771	6.327	2.463	2.1	4.595	3.091	3.002	2.902
G13	12.07	9.72	7.375	-4.079	7.549	-	6.177	3.547	3.088	5.034	4.9	4.532	4.215
G14	11.9	9.297	6.87	-4.394	7.876	-	6.283	2.821	2.514	5.218	4.427	4.391	
G15	-	-	-	-5.657	8.018	-	6.284	3.102	2.517	5.093	4.433	4.262	
T16	9.834	-	-	-5.623	7.141-	1.118	6.042	2.711	2.086	4.99	4.45	4.404	4.148
G17	11.56	9.615	7.065	-4.915	7.408	-	5.952	3.451	2.856	4.998	4.45	4.164	4.099
G18	11.74	-	-	-4.383	7.912	-	6.041	2.256		4.977	4.331	4.129	
G19	10.17	-	-	-4.463	8.014	-	6.092	2.662	2.603	4.831	4.211	4.316	4.172

E. STRUCTURAL STATISTICS TABLES

Table S8. NMR restraints and structural statistics for bundle of eight 2MFT.

NOE-derived distance restraints	Non-exchangeable	Exchangeable
Total	571	122
Intranucleotide NOEs	396	19
Sequential (i, i +1)	154	29
Long-range (i, >i +2)	21	74
Torsion angle restraints	0	
Hydrogen bond restraints[a]	8	
Structural Statistics NOE violations >0.3 Å	0	
Pairwise heavy atom rmsd (Å) all residues	1.06	

Table S9. NMR restraints and structural statistics for a bundle of eight structures formed by the sequence 2M6W.

NOE-derived distance restraints	Non-Exchangeable	Exchangeable
Total	645	379
Intranucleotide NOEs	427	69
Sequential (i, i +1)	175	76
Long-range (i, >i +2)	43	234
Torsion angle restraints	195	
Hydrogen bond restraints[a]	64	
Structural Statistics NOE violations >0.3 Å	0	
Pairwise heavy atom rmsd (Å) all residues	0.664	
Pairwise heavy atom rmsd (Å) structured[a]	0.554	

[a]Residues 1-4, 7-13, 15-18 and 20-24

Table S10. The applied NMR restraints and structural statistics for the bundle of eleven final structures for DNA sequence 5J6U.

NOE-derived distance restraints	Non-exchangeable	Exchangeable
Total	1083	113
Intranucleotide NOEs	621	32
Sequential (i, i +1)	295	28
Long-range (i, >i +2)	20	53
Torsion angle restraints	0	
Hydrogen bond restraints	65	
Structural Statistics NOE violations >0.3 Å	0	
Pairwise heavy atom rmsd (Å) all residues	0.682	

Table S11. NMR restraints for bundle of ten final structures for DNA sequence 5J05.

NOE-derived distance restraints	Non- Exchangeable	Exchangeable
Total	428	94
Intranucleotide NOEs	335	38
Sequential (i, i +1)	78	11
Long-range (i, >i +2)	15	45
Torsion angle restraints	0	
Hydrogen bond restraints[a]	48	
Structural Statistics NOE violations >0.3 Å	0	
Pairwise heavy atom rmsd (Å) all residues	0.994	
Pairwise heavy atom rmsd (Å) structured[a]	0.517	

[a]Residues 4-6 and 10-13.

Table S12. NMR restraints and structural statistics for the ten lowest energy structures formed by DNA sequence 5J4W.

NOE-derived distance restraints	Non-exchangeable	Exchangeable
Total	425	75
Intranucleotide NOEs	286	31
Sequential (i, i +1)	112	19
Long-range (i, >i +2)	27	56
Torsion angle restraints	0	
Hydrogen bond restraints	40	
Structural Statistics NOE violations >0.3 Å	0	
Pairwise heavy atom rmsd (Å) all residues	0.644	

Table S13. NMR restraints and structural statistics for the ten lowest energy structures formed by the DNA sequence 5J4P.

NOE-derived distance restraints	Non-exchangeable	Exchangeable
Total	532	85
Intranucleotide NOEs	361	27
Sequential (i, i +1)	135	16
Long-range (i, >i +2)	36	42
Torsion angle restraints	0	
Hydrogen bond restraints[a]	32	
Structural Statistics NOE violations >0.3 Å	0	
Pairwise heavy atom rmsd (Å) all residues	0.410	

Table S14. NMR restraints and structural statistics for the bundle of ten final structures formed by DNA sequence 2M6V.

NOE-derived distance restraints	Non-Exchangeable	Exchangeable
Total	347	195
Intranucleotide NOEs	226	31
Sequential (i, i +1)	100	44
Long-range (i, >i +2)	21	120
Torsion angle restraints	167	
Hydrogen bond restraints[a]	36	
Structural Statistics NOE violations >0.3 Å	0	
Pairwise heavy atom rmsd (Å) all residues	0.853	
Pairwise heavy atom rmsd (Å) structured[a]	0.517	

[a]Residues 1-2, 4-9, 11, 13-14, and